The Hunt for the Last Respondent

The hunt for the Last Respondent

Nonresponse in sample surveys

De jacht op de laatste respondent

Nonrespons in steekproefonderzoek
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. Dr. W.H. Gispen,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op
19 mei 2005 des middags te 16.15 uur

door
Ineke Anežka Lucia Stoop
geboren op 14 maart 1953
te Den Haag

promotoren: Prof. dr. J.J. Hox, Faculteit Sociale Wetenschappen UU
Prof. dr. J.G. Bethlehem, Faculteit der Econometrische Wetenschappen en Econometrie, Universiteit van Amsterdam
Prof. dr. A.E.Bronner, Faculteit der Maatschappij- en Gedragswetenschappen, Universiteit van Amsterdam

Dit proefschrift werd mogelijk gemaakt met financiële steun van het Sociaal en Cultureel Planbureau.

SCP Report 2005/8
DTP: Mantext, Moerkapelle
Cover Illustration © 1952 Norman Saunders, *Murder by the Pack.* Permission granted by David Saunders
Cover design and flow diagrams: Bureau Stijlzorg, Utrecht
Editing cover illustration: Ien van Laanen, Amsterdam

ISBN 90-377-0215-5
NUR 740

Distribution in the USA & Canada : Transaction Publishers, New Brunswick (USA)
Distribution in Belgium: Maklu-Distributie, Somersstraat 13-15, 2018 Antwerp

Social and Cultural Planning Office (SCP)
Parnassusplein 5
2511 VX The Hague
The Netherlands
Tel. (+31) (070) 340 70 00
Fax (+31) (070) 340 70 44
Website: www.scp.nl
E-mail: info@scp.nl

Contents

Acknowledgements

Except for the conference participant who warned that, “In Alaska people have shotguns and I would definitely not recommend conducting a follow-up survey among persistent refusers there”, I have met almost exclusively with encouragement and support during the writing of this study. The remark was made after a presentation by Frans Louwen and myself at the International Conference on Survey Nonresponse, held in Portland, Oregon in October 1999, in which we outlined our ideas on how to measure nonresponse bias and asked for suggestions on how to proceed. Even this remark from the Alaskan researcher helped in the end, as it suggested the title for this thesis.

The directors of the Social and Cultural Planning Office of the Netherlands, Theo Roes and Paul Schnabel, agreed to spend a substantial sum of money on conducting the follow-up survey on nonresponse and excused me from part of my regular duties to allow me to analyse the data and write this thesis. Even more important was their concern for data quality, their faith in the project and their personal involvement in the follow-up survey. Working on a thesis at the SCP has been a pleasure, with a well-oiled machine in place to take care of lending books and ordering articles, preparing data and providing documentation, checking English and editing the text, drafting figures and designing a cover. An advisory group read drafts of my thesis, saved me from digressions and helped me to stay focused. My colleagues from the department of Data Services and IT were a great help throughout the process as they were eager to know what I was doing, and showed alternately that they could well do without me and appreciated my involvement. I would especially like to thank Leen Both for taking over my duties, listening to my worries, helping me out when my system crashed and devising a tool for the sensitivity analyses reported in chapter 9.

Frans Louwen, with whom I attended the Portland conference, is a director of GfK Panel Services Benelux bv, the organisation responsible for the fieldwork in the surveys analysed in this study. GfK took care of the detailed registration of the fieldwork process and conducted the follow-up survey among nonrespondents. This survey would not have been a success and this study would not have been written without the personal involvement of the GfK staff and the enthusiasm and efforts of the interviewers, who obtained the cooperation of more than 70% of a sample of persistent refusers.

Many people whom I first met at the Portland Conference are regular participants of the International Workshop on Household Survey Nonresponse, a group of dedicated researchers who aim at furthering nonresponse research, sharing ideas and resources and setting a research agenda. Edith de Leeuw introduced me into this workshop and from the first meeting I attended in Budapest in 2000, workshop

participants have showed an interest in the presentations I gave at subsequent meetings – which are all incorporated in this study – and provided ideas for analyses and further study. I would like to thank them all, and especially Edith for introducing me, Mick Couper for looking at the follow-up survey data and Vasja Vehovar for suggesting in Budapest that I should write a book on this and asking at each annual meeting when it would be published.

My friends and colleagues from the European Social Survey (ESS) have played an important role in preventing tunnel vision and sharing their knowledge of survey quality with me. The members of the Central Coordinating Team gave me a crash course in sampling, translation, questionnaire design, measurement error, fieldwork monitoring, data archiving and data dissemination, and in the intricate world of EC calls and proposals. The national coordinators – responsible for the fieldwork in the participating countries – made me acutely aware of cross-cultural differences and the great efforts required to obtain high response rates. The ESS has enriched my life, and being part of the CCT has been a great pleasure. I would like to thank them all, and in particular Jaak Billiet and Michel Philippens and their colleagues from Leuven University for the free exchange of ideas and analyses. We had an exciting time designing standard contact forms, seeing them being implemented all over Europe, meeting criticism and complaints regarding the additional burden but also appreciation and acknowledgement of the wealth of information on fieldwork and the response process they provided.

My supervisors, Joop Hox, Fred Bronner and Jelke Bethlehem, came to this project from diverse spheres and gave me complementary comments and advice, Joop Hox as a social science methodologist supervising (PhD) students in different areas of survey quality and social methodology, Fred Bronner with a background in market research and Jelke Bethlehem with a background in mathematics and statistics. Their support and recommendations were of great help. Jelke Bethlehem and Barry Schouten from Statistics Netherlands analysed data from the follow-up survey in their quest for better weighting models, and involved me in their plans to partially replicate the follow-up survey among nonrespondents in official statistics. Jacqueline Meulman, Paul Schnabel and Peter van der Heijden agreed at short notice to be members of the doctorate committee. I would like to extend my thanks to all of them.

Many other people have been supportive one way or another. H. Russell Bernard from *Field Methods* repeatedly asked me to submit a paper for publication. Parts of this paper (Stoop, 2004a) have been incorporated in this study, in particular the design and evaluation of the follow-up survey in chapter 6. Peter Smeets from Statistics Netherlands helped in finding a legible form in which to depict cooperation across different contacts. David Saunders graciously allowed the use of the cover of *Murder by the Pack*, designed by his father Norman Saunders in 1952, and gave suggestions on how to adapt this cover to the topic of this study.

A doctoral candidate is also supposed to have a private life. Carien Langeveld, Steef de Bie, David and Eva made sure I went on skiing holidays, organized exuberant *Sinterklaas* celebrations and shared pizzas. Beatrijs Ritsema, Maarten Huygen, Felix, Michiel and Saskia took care of animated Christmas dinners. Ko Oudhof made sure we kept on sampling restaurants in The Hague and Leiden. Mark and Elly Meelker took care of me when I had a nasty bicycle accident and showed me how important neighbours are.

A very important person for me, both in general and in this period in particular, is my sister Miranda. I owe her a great deal. She was excited about the work I was doing, invited me regularly to sample her home cooking, and took care of family life involving my father Nol, my brother-in-law Arie, and my niece and nephew Marieke and Ernst. The latter regularly asked when my book would be finished, and if it would be thicker than Harry Potter, and recently, what to wear at the ceremony when I would obtain my degree.

My mother Jopie died in August 2003, and my father is still feeling the loss acutely. It is a comfort for him that my mother knew that I was writing a thesis and that she was proud of me. Even in periods of great grief my father still would ask about my progress. I would like to thank my parents for their care, love and support and for emphasizing the importance of receiving a good education, having a fulfilling job and using your talents.

Ineke Stoop, May 2005

1 Introduction

1.1 Nonresponse: a subject of concern?

High nonresponse rates pose a major threat to survey quality as they can cause unwanted systematic deviations from the true outcome of a survey. Combating nonresponse requires increased fieldwork funds and prolonged fieldwork duration. Even when extended field efforts have resulted in a response rate of 70%, which is generally considered to be quite good, nonresponse bias can occur. The Netherlands is internationally notorious for its low response rates and the problem seems to be getting worse over time (De Leeuw and De Heer, 2002). Therefore nonresponse should be a serious source of anxiety, especially in the Netherlands.

Considering the attention the media pay to the outcomes of surveys and opinion polls it is surprising that the methodology of survey research hardly receives any attention in that same media, except when there is an apparent controversy between survey researchers or when survey outcomes are disputed and fault is found with poor survey quality. Considering the very high costs of survey research it is remarkable that market researchers sometimes express their concerns about the lack of knowledge about or interest in survey methodology of survey sponsors. Considering the importance of survey data for policy making, it is surprising that not responding to governmental surveys is seen more as a sign of independence of spirit than as a lack of civic responsibility. And finally, considering the great importance of empirical data in academic social research it is surprising that survey research methodology itself is an undervalued topic in academic research programmes.

In 1992 the Dutch sociologist Herman Vuijsje discussed the paradox that the Dutch government needs data to pursue the welfare of its citizens whereas these same citizens are increasingly hesitant in providing that information:

> 'Nonresponse is seen as an understandable and commendable warning from well-meaning citizens. 'Distrust of authorities' is observed, but this distrust is considered almost by definition as an expression of social engagement: a protest reaction stemming from a sincere concern about privacy. .... This too is a characteristic paradox of the Dutch welfare state: in a country with one of the most citizen-friendly governments that have ever existed, a broad sense of social distrust is developing of the very same government. This distrust, which hampers statistical data collection for policy purposes, is respected and honoured by government rather than contested. In some circles all but an enemy image has grown of governmental, research and medical organisations. This image expresses itself rather unilaterally: citizens do not turn away from the

government in its function of provider of subsidies or medical services, but they do when the government asks for the information required for planning these services.' (p. 56, translated from Dutch)

This comes rather close to the Canadian MP who said in 1980: 'Statistics Canada surveys have come under attack from people all across the country, for very good reasons – they constitute a blatant and legalized invasion of privacy ...' (cited by Goyder, 1987, p. 26)

The general attitude towards survey participation appeared to be captured nicely by the fact that Dutch respondents reacted with amusement and incomprehension to a statement adopted from a US study 'I think it is my civic duty as a citizen to agree to participate when asked in most surveys that come from a government organization' (Couper and De Leeuw, 2003, p. 170). A final illustration of the ambiguous survey climate in the Netherlands is provided by the content of www.uwmeningtelt.nl, the information centre for market and opinion research. In the recent past Uw Mening Telt (UMT) ran a radio commercial *Doe je voordeel met je oordeel* (Take Advantage of Your Opinion) encouraging the general public to participate in survey research. The UMT site in summer 2004 was purely and simply an opt-out facility for unannounced telephone and Internet surveys.[1] If market research information centres make it their main concern to protect Dutch citizens from annoying surveys, it makes sense that these very same citizens do not feel any obligation to participate.

In the mid 1990s, the Dutch political scientist Gerard Visscher (1995) questioned the validity of the Dutch Parliamentary Election Study and assumed that the declining response rate in this survey most likely reflected a decreasing interest in politics, whereas political scientists generally assumed that the stable interest in politics among the respondents reflected the attitude of the entire Dutch population. Visscher's publication instigated a discussion about nonresponse in the media and among policymakers, politicians and the general public. Headlines in major Dutch newspapers contested the value and the quality of survey research: *Refusers make life hard for researchers* (Trouw, 12 September, 1996), *Sovereign remedy against interviewers: ask for money* (Volkskrant, 26 April 1997), *Nonresponse worst culprit* (Trouw, 8 October, 1997), *The Dutch suffer from interview fatigue* (regional newspapers, 20 November 1997) and *The worst superstition is the belief in survey sampling* (referring to a study on religious affiliation, Trouw, 28 June, 1998). Nonresponse in surveys from Statistics Nether-

1 Luckily an exception could be made for surveys mounted by Statistics Netherlands and the government, surveys among clients, very short surveys (less than 5 minutes) and surveys that provide a monetary incentive. All members of the Dutch market research association (www.marktonderzoekassociatie.nl) are required to adhere to this opt-out option (MOA-nieuws, 2003). It seems similar to the Do not call register in the US hich, however, explicitly excludes market research (ESOMAR, 2003b; Bowers, 2004).

lands and possible bias due to refusal were even discussed in Parliament (Tweede Kamer, 1996). Visscher's philippic against the election study and survey research in general led to a flood of studies on nonresponse and political interest, and heated discussions among Dutch political scientists which have continued to this day on whether or not declining response rates indicate a declining interest in politics (see for instance Visscher, 1995; Smeets, 1995; Voogt, Saris and Niemöller, 1998; Aarts and Van der Kolk, 1999; Van der Kolk, 2001; Dekker, 2002; Van Goor, 2003; Voogt, 2004). The reason for this interest might be, as Smeets (1995, p. 308) points out, that in election surveys possible nonresponse bias is extremely visible as voter turnout figures can be compared to official election turnout figures. In other areas, however, interest in nonresponse and survey quality is far less pronounced.

In the Dutch journal *Sociologische Gids* which regularly contains contributions on nonresponse, Ellemers (2004) calls the 'nonresponse scandal' the biggest problem in Dutch social sciences and a major impediment to obtaining knowledge on Dutch society. He feels that for years the possible presence of nonresponse bias has been ignored or played down even in surveys with a nonresponse rate of only 20% or 30%, and worse, that hardly any serious attention has been paid to the question of how this fault can be remedied. The present study can be seen as an ultimate form of 'muckraking': it aims at divulging the gory details of the nonresponse scandal, assessing the gravity of the situation and presenting recommendations to cope with nonresponse.

1.2 Survey research: the cast

Several stakeholders are involved in survey research and play a role in commissioning fieldwork, conducting surveys, interviewing respondents, analysing and interpreting data, using survey results and improving data collection methods. For each of these nonresponse is or should be important, if only because nonresponse increases fieldwork costs. In her book on the history of surveys Converse (1987) describes their multiple roots, namely in scientific research conducted by universities and commercial research conducted by market research organizations. This section will proceed with a very brief overview of the role of universities, market research organizations, Statistics Netherlands and other parties involved in surveys and nonresponse in the Netherlands.

Within academia, survey research does not seem to have a very prominent position. Morton-Williams (1993, p.2) states in her introduction to a comprehensive book on survey interviewing that '... In Britain, the failure of universities to become seriously involved in survey research has been a factor in the slower development of survey professionalism in this country so that higher status still tends to be accorded to the subject specialist than to the survey specialist, who is sometimes regarded simply as a survey facility manager rather than as an expert in survey and questionnaire design, analysis and interpretation'. In the Netherlands the situation is comparable,

and interest in surveys even seems to have declined over the years. Ten years ago, the Social Sciences Council, an advisory Council of the Royal Netherlands Academy of Arts and Sciences (Kalfs and Kool, 1994) asked fifty senior survey researchers about their experiences with nonresponse. A quarter of these researchers felt that nonresponse caused substantial bias, a third that nonresponse had no perceptible effect on survey quality, two-thirds that not enough was known about nonresponse, and a large majority that more research was needed. This latter recommendation has not been translated into more systematic attention to survey methodology and nonresponse research. Among Dutch social methodologists are a number of international acknowledged experts in questionnaire design and nonresponse, but nowhere in the scientific programmes of the Netherlands Organisation for Scientific Research (NWO) attention is paid to nonresponse or to survey research or to social research methodology in general. Research methodology is not the most popular discipline in the social sciences in the Netherlands, survey research seems even less attractive.[2] Morton-Williams (1993, p. 7) quotes Gorden (1975) 'There is an unfortunate tendency to consider data gathering as a lower status activity than data analysis or theorising about the data'.

Market research organizations are the second major player in the area of social surveys. In the past, commercial market organizations set up special studies of survey participation and nonresponse. In 1977 and 1978, Van Tulder analysed the reactions of surveys participants when they were asked to cooperate in a face-to-face survey (section 3.3, 3.5). Also in 1978, the market research group Inter/view-groep (1978) published a study on nonresponse aimed at generating a discussion on response enhancement procedures and at greater transparency in nonresponse reporting. At that time a response rate of 95% was feasible. Since then, the yearbooks of the association of Dutch market research organizations[3] have regularly paid attention to developments in nonresponse in household surveys. De Leeuw (1997) gives an overview of earlier work, more recent contributions have been made by De Kruijk and Hermans, 1998; Pondman, 1999; Van der Kolk, 2001; Kalfs, 2001; Carton, 2001; Stoop, 2002; Deutskens et al, 2003. As recent as June 2000, UMT organized a success-

2 Market research as a career seems to suffer from a similarly low appeal. 'Some feel the perceived low level of professionalism, and the resulting limited career-perspectives, make the profession a relatively undesirable place for high level talent. .... MR people have the image of a grey mouse, a dry book-keeper. MR is regarded as a transition to higher positions. The products of MR are tables and statistics, not Ferrari's or Armani' (Heeg, 2004).

3 The current association of market research organizations is the MarktOnderzoekAssociatie www.marktonderzoekassociatie.nl (MOA). The MOA was founded in 2000 and is the result of a merger of the Nederlandse Vereniging van Marktonderzoek en Informatietechnologie (NVMI), the Vereniging van Marktonderzoek-organisaties (VMO) en Uw Mening Telt (UMT), the PR organization promoting participation in market research. A precursor was the Nederlandse Vereniging van Marktonderzoekers.

ful meeting on nonresponse with national and international speakers that even drew the attention of the national press (Van Delft, 1999). Nowadays, low response rates and high costs of random sample face-to-face surveys have prompted all leading market research organizations in the Netherlands to look for alternatives (ESOMAR, 2003a). Telephone recruitment of respondents, access panels and online research are gaining popularity. As a consequence, random sample face-to-face surveys are turning into something of an anomaly or, formulated positively, into a high-premium niche. A cause, or a symptom of this development is that interviewers increasingly appear to dislike the recruitment phase of surveys. Administering a CAPI interview is considered a professional, high-status activity, whereas trying to obtain the cooperation of selected sample households is considered to be a low-status, unrewarding chore. This is not an entirely new perspective: 'The overarching role that the interviewer wants to establish is that of 'survey interviewer': a professional and impartial collector of important information for a serious and interesting purpose. But initially a subsidiary aspect of that role is as a petitioner making a request' (Morton-Williams, 1993, p. 124).

The third major party in Dutch survey research is Statistics Netherlands (CBS) which conducted an expenditure survey as long ago as the 1930s, although surveys started to become popular only in the 1970s. In the last decades of the twentieth century, surveys lay at the base of a wide range of official statistics.[4] According to De Heer and De Leeuw (1999, p. 53) Statistics Netherlands was originally a follower rather than a leader in the area of interview surveys. Only during the late 1980s did it start to set to the tone and developed Blaise, now one of the leading CAPI programs, introduced coordinated questions across surveys and set up a cognitive testing laboratory. In recent decades Statistics Netherlands has collected survey data on labour, housing, health, environmental behaviour, political preferences, working conditions, voluntary work, victimization, etc. At present an increasing amount of data from administrations and public records becomes available that are integrated in a huge virtual databank.

Statistics Netherlands has been struggling for years with high nonresponse rates in household surveys. Nonresponse rates of 45% in face-to-face surveys were common. Te Riele (2002) demonstrated in an analysis of a wide range of statistical surveys that these low response rates will inevitably lead to bias. Statistics Netherlands tries to combat this situation in several ways. Firstly, as mentioned above, it increasingly replaces survey data with data from administrative records and public registers. Secondly, it uses information from these records to estimate bias and correct for nonresponse. Thirdly, it has recently organized its entire fieldwork organization and introduced, among other things, regional managers who have to closely monitor the

4 In their treatise on the quality of official statistics titled 'Can a Statistician deliver' Platek and Särndal (2001) even seem to equate statistical quality with survey quality.

progress and quality of fieldwork. This close monitoring and regional management may be one of the reasons for a recent increase in response rates. Another reason may be the recent introduction of a strategic programme aimed at improving response rates, coverage and procedures to correct for nonresponse.

The Social and Cultural Planning Office of the Netherlands(SCP) is a prolific user of survey data from Statistics Netherlands and other providers. Over the years it has commissioned a wide range of surveys among specific groups or on topics on which the SCP working programme required data that were not available from existing sources. As a consequence, the SCP has willy-nilly grown into a major sponsor of large-scale semi-official social surveys in the Netherlands (Stoop, 2000) and has become increasingly involved in survey research methodology, partly because empirical data are the cornerstone of its substantive research and partly because commissioning fieldwork requires methodological expertise in drafting specifications, assessing proposals and balancing the trade-off between costs, timeliness and survey quality. This need for methodological expertise has been exacerbated by an increase of both data needs and survey costs. Low response rates became a rather urgent problem in 1991 when the response rate of the four-yearly face-to-face Amenities and Services Utilization Survey (AVO) stuck at a measly 43 percent instead of the mid 60 percent of earlier rounds. Since that time SCP has been involved in nonresponse research with higher response rates and this study as a result.

The protagonists of survey research are of course the people out there, the foot soldiers (interviewers), the quarry (respondents) and the getaways (nonrespondents). There is an unfortunate tendency to ignore the importance of interviewers. Interviewing is a poorly paid, low-status job. In the Netherlands until recently interviewers earned approximately the same as a cleaner. Now their formal position has improved as a consequence of new tax and labour laws. This somewhat menial position is not to be blamed on market research organizations. Funders of fieldwork would be rather unpleasantly surprised if the costs of interviewers were doubled to secure better pay, a better position and better training. Goyder (1987, p. 30) wryly called survey interviewers '...those forgotten plebeians who actually perform interviews...' Converse and Schumann (1974, pp. 33-36) give a poignant insight into their marginal position: 'Our own interviewers report that the Detroit Area Study staff sometimes display utter lack of interest in the human challenges of interviewing in their single-minded concentration on clerical competence'. This lack of interest and poor remuneration is especially worrisome as it is generally acknowledged that high interviewer morale and intrinsic interviewer motivation are decisive factors in obtaining high response rates (Morton-Williams, 1993, pp. 65-66; Lemay and Durand, 2002, p. 35). This seems to imply that survey sponsors and market research staff should give interviewers tangible evidence that their contribution is important.

The respondents are the second group in the field. The question as to why people respond induced Goyder (1987, preface) to write his important monograph on nonresponse. What does not generally come up in the sophisticated studies on survey cooperation to be discussed later, is that respondents are often treated abominably. They are 'sugged' (Selling Under the Guise of Research), receive advance letters but no interview, are deluded with the false prospect of a very short interview, receive no proper information on the purpose of the survey, wait for interviewers who do not keep their appointments and have to answer impossibly long questionnaires with impossibly stupid and irrelevant questions. Of course this does not apply to the good surveys to be discussed here, but still, one sometimes wonders why anybody would ever take part in a survey. This leads to the most interesting group of all, the nonrespondents. Groves and Couper (1998, p. 62) worry about the high response rates in the surveys in their meta-analysis (82% to 97%) and the even higher cooperation rate (87% to 98%). If overall survey cooperation is more than 90%, it is difficult to detect subgroups that exhibit large differences in cooperation. Luckily, nonrespondents and refusers in the Netherlands are more tenacious. If not for the high nonresponse rate and the high refusal rate – and the high cooperation rate in a follow-up survey among refusers – this study would not have been written.

All parties involved in survey research are aware of the sad truth that in general funds for data collection are limited and that in designing a survey it will not always be clear which allocation of funds will do most to improve accuracy: 'For example, a survey designer may ask, 'Will nonresponse bias be reduced more by increasing the interviewer training budget by 10,000 USD or by allocating these funds to the nonresponse follow-up operation or by using them to increase the amount of incentives paid to respondents?' (Biemer, 2001, p. 26). As long ago as 1987 Converse (p. 411-412) placed survey research in the perspective of *The Relentless Need for Big Money*. Funding plays a major role in Converse's book, as it does in Groves' 1989 *Survey errors and survey costs*. As Platek and Särndal say (2001, p. 6): 'The discriminating user ... is aware that poor quality may have its roots in underestimated total cost of the survey, insufficient control of the survey operation, or improper allocation of cost to operations'. High-quality survey data are necessary for both academic and policy purposes, and data collection is expensive, face-to-face surveys are very expensive and good face-to-face household surveys are seriously expensive. Therefore, all parties involved should be aware that vast ambitions cannot be financed out of petty cash. This study aims at conveying the message that data collection is part of a scientific process and should not be a black hole in between a careful formulation of research questions and meticulous analysis of the data. This scientific approach will require extra funds for fieldwork monitoring and quality control, activities the benefits of which will not always be immediately obvious to substantive researchers and survey sponsors.

Related to the need for money is the need for a professional well-run organization. Chapter 4 will show that there is extensive literature on issues such as advance letters and incentives. Whereas these can have noticeable effects on nonresponse rates, organizational factors can have very large effects. If interviewers are hard to find in

particular regions, if interviewers are badly paid and are not motivated to call again after noncontacts or to re-approach unfriendly refusals, if interviewers drop out in some regions and this is not observed early on, if the workload of interviewers or the fieldwork organizations is too large and everybody has to move on to the next survey, response rates will suffer massively.

And finally, in addition to money and a professional organization, surveys do need tender loving care. One of the brochures on survey quality from the American Statistical Association (2000) is titled Judging the Quality of a Survey. In trying to answer the question 'How Good Is a Particular Survey?' the authors come up with the following answer:

> 'The potential for problems is a reality in all surveys today. The good news is, however, that researchers have found at least partially effective ways to deal with most problems that occur. The main issue for the discriminating user of results from any survey is to determine whether problems like those described previously were recognized, and steps were thoughtfully taken to deal with them. Indeed, the quality of a survey is best judged not by its size, scope, or prominence, but by how much attention is given to dealing with all the many important problems that can arise.'

1.3 *Perspectives on nonresponse research: questions to be addressed*

This study is rooted in a firm belief in the superior quality of classical survey methodology. This implies neither that, for instance, face-to-face interviewing is believed to be necessarily superior to online interviewing, nor that nonresponse is the only problem. The focus on face-to-face surveys is dictated by practical considerations, the focus is on nonresponse because this is generally acknowledged to be a serious source of bias, and the embedding in classical survey methodology seems to be the best approach to improve total survey quality (see chapter 2). The focus on nonresponse opens three perspectives. In the brochure *Judging the Quality of a Survey* from the American Statistical Association (2000) a distinction is made between those who aim at enhancing response rates, those who want to measure the effect of nonresponse on survey outcomes, and those who want to adjust the survey outcomes in order to compensate for any effects of nonresponse. Groves (1989, pp. 283-288) describes a hilarious conversation reflecting the inability of describers and modellers to understand each other's point of view.

The enhancers tend to focus on fieldwork strategies that minimize nonresponse rates by ensuring that sample units are contacted and promoting survey cooperation. In doing this they generally acknowledge that response rates differ among different groups in society and that a range of response enhancing measures should be developed to cater to the needs and preferences of these different groups. Measurers accept response rates

more or less as they are, and assess the effect of nonresponse on survey outcomes. In doing this, they compare characteristics of respondents and nonrespondents, discriminate between noncontacts and refusals, and hard-to-contacts and temporary refusals, and estimate the effect of nonresponse using information from the sample frame and interviewer observations. Modellers or adjusters (also fondly called the Greek people, in the International Workshop on Household Survey Nonresponse, because they use a lot of mathematical formulae) tend to think in terms of total error and bias. They model nonresponse and adjust for nonresponse by weighting using auxiliary information from the sample frame, fieldwork behaviour and interviewer information.

Nonresponse research based on these differences might show counter-purposes. Two examples can illustrate this. Interviewers may gear their timing of calls to previous knowledge on when the sample person is likely to be at home, which is likely to increase response rates. Measurers might prefer calling on respondents to be based on standard call schedules, whereas adjusters would like to have information on the presence and type of this earlier information. Secondly, response enhancement techniques that cater to the needs of individual respondents in a flexible way may complicate the modelling of the relationship between survey variables and response propensities. Whereas enhancers might try to work out who might most easily be persuaded to cooperate after an initial refusal and direct their efforts at those promising cases, measurers might prefer re-approaching refusals and assessing their willingness or refusal, whereas adjuster would need transparent information on interviewer behaviour, doorstep interaction and fieldwork decisions as additional variables in their models. Of course, these examples exaggerate the differences, and enhancers, measurers and adjusters could and should profit from each others work. Enhancers could optimize their fieldwork strategies based on the research outcomes of measures, adjusters should try to find out what goes on in the field in order to know what could and should be modelled. Lynn (1996) distinguishes two approaches, namely enhancing response rates and adjusting for nonresponse in the analysis stage and recommends combining them. The present study aims at combining the perspectives of an enhancer and a measurer, but it thus not extend to actually adjusting for nonresponse.[5] The following issues will be central.

Noncontact and noncooperation: different dimensions of nonresponse

Response enhancement in face-to-face surveys concerns two separate field processes, namely achieving contact and obtaining cooperation. If no contact has been established nonresponse is a certain outcome, and if contact has been established the prospective respondents may still refuse. The distinction between contacting

5 The results of the empirical study presented in chapter 6 to 9 could be used to adjust for nonresponse. This has been tried in Schouten, Bethlehem and Stoop (2002) without much success as the relationship between auxiliary variables in the study for which population totals were available and some of the target variables was too small to satisfactorily correct for selective non-response.

sample persons and obtaining cooperation is essential for different reasons. Firstly, environmental, household and individual characteristics of the prospective respondent may not be under the researcher's control but may differentially influence contact and cooperation. Knowledge on who is hard to reach and who is reluctant to cooperate is therefore of vital importance when setting up a survey, devising response-enhancing measures and allocating funds effectively, which is why the first part of this study comprises an inventory of the literature on who is hard to reach and less likely to cooperate. In measuring and adjusting for nonresponse bias the ease of contact and reluctance to cooperate of respondents are generally considered to be important auxiliary variables which should clearly be distinguished as bias due to noncontact may differ from bias due to refusal. For this reason the distinction between contact and cooperation and the development of instruments to measure accessibility and amenability are central in this study.

Response enhancement

Survey cooperation depends not only on the prospective respondent, but also on the survey in question and the interaction with interviewers. A number of these issues should be treated as given. Some survey topics, for instance, may be more popular among respondents than others. Response enhancers cannot change the topic in order to obtain a higher response rate. In order to obtain high response rates they might try to convey that this topic is of general interest, however, or include extra incentives to compensate for unattractive survey topics. Measurers and adjusters include information on the topic in their analyses and models. Other issues may be influenced in order to obtain high response rates, such as survey mode, advance letters, incentives, interviewer experience and fieldwork duration. This study will give an overview of response enhancing strategies and also present two Dutch surveys in which these strategies resulted in quite acceptable response rates, even from a comparative point of view.

Bias reduction

As mentioned above, this study originates from the concern that low response rates affect the accuracy of survey outcomes and can cause unwanted systematic deviations from the true outcome. It aims at actually measuring the increase in accuracy obtained through enhanced response rates in a particular survey and assessing the nonresponse bias that remains after extended fieldwork and increased response rates. Put differently, the main question to be addressed is whether enhancing response rates does result in outcomes that differ from those of a low-effort survey, and in outcomes that are closer to the true value. This question is generally outside the scope of response enhancers and should be seen from a measurer's perspective. Addressing this question requires information on the way in which additional respondents were obtained, on their background characteristics and on their survey outcomes. Furthermore, it requires independent information on final nonrespondents. The next step would be adjusting for nonresponse. The auxiliary informa-

tion collected on the response process and the additional respondents could help modellers in weighting for nonresponse and, if enhancing response rates reduces bias, nonresponse adjusters could use 'difficult' respondents as proxies for final nonrespondents.

Questions of a measurer

The different perspectives on nonresponse research distinguish those who believe survey quality can be improved, or nonresponse bias reduced, by either enhancing response rates or adjusting for nonresponse. The measurement perspective used in this study aims at collecting and analysing information with which this question can be answered. Based on the nonresponse literature, an analysis of nonresponse in two surveys and an experimental study, six questions will be addressed.

This study has been inspired by concerns about the consequences of nonresponse for survey outcomes. The first question aims at specifying when and why nonresponse can cause nonresponse error:

Q1 *Why are high nonresponse rates a reason for concern?*

A more practical question is which groups in society are less likely to respond, either due to noncontact or to noncooperation. The probability of response can be related to socio-demographic, psychological and survey-specific factors:

Q2 *Who are less likely to respond, either because they are more difficult to contact or because they are more reluctant to cooperate?*

A second practical question is inspired by the assumption that nonresponse error is likely to be smaller when the response rate is higher:

Q3 *How can response rates be enhanced?*

By definition, nonresponse is difficult to study because little information is generally available about nonrespondents. Different sources and tools can provide a part of the missing information, and nonresponse models can specify the relationship between field efforts, types of nonresponse and nonresponse error. Appropriate tools and models will be investigated as part of the question:

Q4 *How to study nonresponse?*

Whereas the first question indicated a general and theoretical concern about nonresponse, and the third question simply assumed that higher response rates would minimise nonresponse error, this study also examines whether this concern will always be justified and whether higher response rates will always lead to better data quality:

Q5 *Do enhanced response rates improve the accuracy of survey outcomes?*

When these questions have been answered the question remains as to how nonresponse error can be minimised and whether enhancing response rates is the best way of doing this:

Q6 How to combat nonresponse error and allocate funds effectively?

1.4 Outline of this study

Nonresponse is a well-researched issue and there is quite a body of literature available. Individual studies are often focused on single issues; the response process is rarely closely monitored and reported and independent evidence on bias is infrequently available. For this reason, the questions presented above will be addressed in two phases. Part 1 of this study gives an overview of the literature on nonresponse. This will be followed by the presentation and analyses of two surveys in which many of the recommendations for response enhancement and process measurement from the literature have been implemented. After the overview of the literature answers to the above questions will be given and research questions for the second, empirical part of this study will be presented. The lay-out of this study described below shows that the questions above will be addressed from different perspectives and on different levels, and thus in different chapters.

Chapter 2 gives an introduction into nonresponse research, explores when and why nonresponse can cause bias, shows how to compute response rates and embeds obtaining response in the framework of random sampling. It will give a short overview of the relationship between interview modes and response rates and then present the two major factors in obtaining response in face-to-face surveys, namely establishing contact and obtaining cooperation, the latter including refusal conversion, i.e. re-contacting initially reluctant respondents in order to obtain their cooperation. Separate models for contactability and cooperation will be introduced, reasons for why it is important to distinguish between contactability and cooperation and for why this is so often not done in nonresponse research. This chapter mainly addresses the question of why high nonresponse rates are a question of concern (Q1) and how to measure contactability and reluctance, which is one factor in studying nonresponse (Q4).

Chapter 3 then proceeds to look at question Q2 and at distinguishing who are generally identified as being hard to contact and reluctant to cooperate or inclined to refuse. This chapter will focus on the socio-structural correlates of nonresponse, more commonly known as background variables. Knowing who is likely not to respond may not suffice when trying to enhance response rates nor to adjust for nonresponse. Background variables may reflect causes why persons refuse but they are rarely the cause itself. Knowledge on causes of nonresponse can help in enhancing response rates and identifying bias. Therefore chapter 4 will delve more deeply into the question of who are inclined to not respond and firstly identify social participation and social

involvement as underlying causes of survey participation – which might be a serious cause of nonresponse bias – and present results of studies that directly measure attitudes towards surveys. If survey attitudes exist that are independent of background variables, information on these attitudes from both respondents and nonrespondents would be valuable in adjusting for nonresponse. In addition, it will be shown that survey participation is not solely a function of demographic, social and psychological characteristics of the prospective respondents, but also of the design of the survey, interviewer' characteristics and respondent-interviewer interaction. The overview in chapter 4 of response enhancement strategies should therefore address Q3 but also show which survey characteristics may affect response rates, how different types of reluctance can be overcome by which strategies and why different types of reluctance may cause bias. The final section of this chapter looks back on the literature, theories and studies presented from a moral point of view and pits behavioural theories in which sample cases are manipulated and cajoled into saying yes to the interview request, against the social exchange view in which responsible citizens are free to decide whether or not the want to cooperate in a survey or not. The question will be addressed of whether it is ethical to try to enhance response rates when the prospective respondents are obviously less than eager to cooperate.

Chapter 5 moves away from the perspective of the response enhancer and investigates approaches to studying nonresponse (Q4). Central in many studies is the assumption that respondents can provide information on nonrespondents either by classifying them according to difficulty or to type. Two models, originally developed by Lin and Schaeffer (1995) will be presented and an overview will be given of empirical studies in which these models have been tested. In addition, tools will be presented to collect information on nonrespondents and therefore obtain independent information to assess nonresponse bias (Q4) and the central assumption will be reviewed that enhanced response rates will improve survey results (Q5).

At the end of chapter 5, in section 5.5, the evidence from the literature will be used to answer the questions above. Furthermore, specific research questions will be formulated that will be answered in the empirical questions of this study, including the final question (Q6). Section 5.5 can be seen of an inventory of what is known and an overview of which questions remain that lack a final answer and that will be addressed in the second part of this study.

Part 2 starts with an overview of the history, content and design of the Amenities and Services Utilization Survey (AVO, in Dutch) 1999 that plays a pivotal role in this study. Chapter 6 explains how the fieldwork was carried out and which problems arose both as a background to the following chapters and to show how information on the fieldwork process was collected. This chapter also pays attention to a central source of information in this study, namely the follow-up survey among persistent refusals to the AVO, and describes which response enhancement strategies were used and why

this follow-up survey among this very difficult group managed to achieve a very high response rates. This chapter gives practical answers to the question of how response rates can be enhanced (Q3), shows how nonresponse can be studied by measuring contactability and cooperation and by obtaining information from nonrespondents through a follow-up survey (Q4).

Chapter 7 focuses exclusively on contacting respondents. It explores the inherent problems in measuring contactability (Q4), investigates who are hard to contact (Q2), how the contact rate can be enhanced (Q3), whether there is a continuum of resistance with regard to contactability (Q4) and whether a high contact rate improves survey quality (Q5).

Chapter 8 proceeds along similar line with respect to reluctance to cooperate and shows how willingness to cooperate can be measured (Q4), identifies those who are immediately willing to cooperate and those who initially refuse (Q2), how refusal conversion can secure a sizeable number of additional respondents (Q3), and investigates if there are different classes of participants (Q4) and if refusal conversion improves survey quality (Q5). This chapter also investigates whether exerting some pressure on sample persons to cooperate leads to poorer data quality.

Chapter 9 goes a step further than most nonresponse studies as it analyses the proceeds from the follow-up survey among nonrespondents and is thus able to provide an answer to the question of whether the extended fieldwork in the regular survey, and in particular the attempts to convert refusals, did improve the accuracy of the survey outcomes (Q5). This chapter will show that a follow-up survey among refusals is a feasible way of studying nonresponse and measuring nonresponse bias (Q4). It will pit a follow-up survey among nonrespondents as a direct measure of nonresponse bias against endeavours to enhance response rates as a means to minimize nonresponse bias and improve survey quality (Q6).

Chapter 10 comprises an excursion to another survey, the European Social Survey, in which high response rates were pursued and achieved, and the response process was recorded in similar ways as in the AVO. The results of the ESS can help to assess the value of the AVO-study, especially as results from many European countries are available and as the topic of the survey is completely different (opinion rather than factual questions). The chapter aims at answering two questions, namely if the nonresponse results of the AVO were particular to this survey and if the instruments developed in the AVO with which to measure contactability and reluctance are feasible in other surveys .

Chapter 11 will look back on the results, discuss the limitations of the present study and come up with recommendations for survey fieldwork and nonresponse research. The study will end with conclusions and a summary in English and in Dutch.

This study covers a wide range of aspects of survey and nonresponse research, but many issues are beyond its scope. It covers household surveys only, not business surveys. It focuses on nonresponse in random sample surveys and pays only passing attention to non-probability sampling. Its main focus is on face-to-face surveys, and not on telephone, mail and Internet surveys, although the other modes will be mentioned where relevant, for instance when telephone recruitment for face-to-face surveys is discussed (section 2.5). It is also mainly concerned with unit nonresponse, and not with other survey quality criteria, such as noncoverage and measurement error. Item nonresponse will be mentioned shortly in section 8.3.3, which investigates whether pushing respondents to cooperate may threaten survey quality and enhance item nonresponse. It will not present formal models for estimating bias, nor will it present procedures to weight for nonresponse. As the empirical data in this study are mainly from a Dutch survey, special attention will be paid to survey research and nonresponse analyses in the Netherlands. The overview of the literature mostly reflects the situation in the US, the United Kingdom, Germany, Belgium, Finland, Sweden and the Netherlands.

Part 1

Nonresponse in sample surveys: an overview of the literature

2 What is nonresponse and why be concerned

2.1 The importance of response rates

Concerns about the quality of survey research and nonresponse are widespread. International journals publish on survey research and methodological issues, international groups of researchers discuss methodological problems and develop shared research agendas, and cross-cultural surveys initiate methodological research and a comparison of methods and results. Table 2.1 gives an overview of these networks, sources and surveys that have been of great importance in writing this book (for a complete overview see the references). Groves' 1989 600 pages India-paper monograph on Survey Errors and Survey Costs looks somewhat forbidding, but contains everything you always wanted (or did not want to know) about survey research. Fowler (2002) manages to discuss all dimensions of survey research in a slim booklet the completeness of which you only realize after having worked in survey research for quite some time. De Vaus (2002) collected papers on all possible aspects of surveys in a four-volume edition. A recent addition is Biemer and Lyberg's (2003) comprehensive overview of survey quality issues. Finally, there is probably no study published on nonresponse in the last five years that does not refer to Groves and Couper (1998) and their models on nonresponse and survey participation. This is indeed the bible of nonresponse research.

In a recent conference on survey quality, the response rate of a survey was generally considered to be a major – or *the* major – quality criterion (Marker, 2002). It is therefore not surprising that the question 'What is an acceptable response rate?' is frequently raised. In the literature circulates the Babbie response rate rule (Babbie, 1992, p. 267) '50% response is 'adequate', 60% is 'good', and 70% is 'very good'. Fowler (2002, p. 42) is less clear: 'There is no agreed-upon standard for a minimum acceptable response rate. The Office of Management and Budget of the federal government, which reviews surveys done under contract to the government, generally asks that procedures be likely to yield a response rate in excess of 75%. In the United States, academic survey organizations often are able to achieve response rates for designated adults in the 75% range with general household samples.' Voogt (2004) concludes from his nonresponse studies that relationships between variables seem rather impervious to nonresponse bias, provided that the response rate has reached an acceptable level (at least 70%). Others set much lower standards. Deutskens et al. (2003, pp. 54-55) considered a response rate of 31% in a web survey promising for the future of online market research (p. 54) and concluded: 'With the achieved response rate of almost 25% is seems possible to administer questionnaires of considerable length via the Internet.'

Table 2.1 International journals and networks important for survey methodology and nonresponse research

journals		
Public Opinion Quarterly (POQ)	POQ	www.journals.uchicago.edu/POQ/home.html
International Journal of Public Opinion Research	IJPOR	ijpor.oupjournals.org
Journal of Official Statistics (JOS)	JOS	www.jos.nu
Field Methods		www.acadimage.com/Field_Methods
Bulletin of Sociological Methodology and newsletter	BMS	www.iresco.fr/bms031119/Welcome-new3.html
networks		
International Workshop on Household Survey Nonresponse		nonresponse.stat.ucla.edu
International Workshop on Comparative Survey Design and Implementation, initiated by ZUMA	CSDI	www.gesis.org/en/zuma/index.htm
cross-national surveys that initiated methodological research		
International Adult Literacy Survey	IALS	www.statistics.gov.uk/ssd/surveys/european_adult_literacy_review_survey.asp
International Social Survey Programme	ISSP	www.issp.org
European Social Survey	ESS	www.europeansocialsurvey.org
benchmark literature (see references)		
Jean M. Converse	1987	Survey Research in the United States
Robert M. Groves	1989	Survey Costs and Survey Errors
Robert M. Groves and Mick P. Couper	1998	Nonresponse in Household Interview Surveys
Floyd J. Fowler Jr.	2002	Survey Research Methods
Paul P. Biemer and Lars E. Lyberg	2003	Introduction to Survey Quality
Jaarboek Marktonderzoekassociatie	MOA	Ontwikkelingen in het marktonderzoek
David de Vaus	2002	Social Surveys. Sage Benchmarks in Social Research Methods (4 Vols.)
American Statistical Association, ASA series	2000	What is a Survey? www.amstat.org/sections/srms (last modification 21 May 2000)
Leadership Group (LEG) on Quality aimed at mproving the European Statistical System (Lyberg et al.)	2001	amrads.jrc.it/WPs%20pages/Quality/Documents/LEGsummary.pdf

Single-mindedly focusing on response rates alone should be discouraged, however (Groves, 1989, p. 133 and 147). Platek and Särndal (2001, p.11) discuss the misunderstandings regarding response rates in a special issue of the Journal of Official Statistics on the quality of official statistics, or rather surveys conducted by statistical agencies: 'All users view high nonresponse as one of the main threats to data quality. But in assessing the effect of a stated rate of 36 percent, they risk falling back on stereotyped ideas, found perhaps in well-meaning texts, or heard from colleagues, of the type 'a nonresponse rate of more than 30 percent (or a similar rule of thumb) will render the survey data useless' ('and they may think this irrespective of the survey conditions'). Biemer and Lyberg (2003, p. 95) express a similar opinion in a treatise on why nonresponse should not be taken lightly: '... nonresponse is a quality

feature that many survey users and sponsors have heard of, and they know intuitively that a high nonresponse rate is not in line with good survey practices. Data users may know very little about how the data were collected and all the steps that have been taken to improve data quality. Therefore, the survey response rate is viewed as an indicator of the quality of the entire survey process.... Indeed, to these users, the nonresponse rate may also be indicative of the competence of a survey organization. High response rates become synonymous with efficient, high-quality data collection operations. So to stay in business, there is a practical reason to keep nonresponse rates low. In the long run, however, the survey community should vigorously educate sponsors and users regarding the totality of survey quality indicators.'

Rather than set store by the response rates *per se*, one should be willing to find out about the mechanism behind nonresponse. Groves (1989) shows that theoretically higher response rates can even result in worse survey estimates when those initial nonrespondents who are converted to respondents are very atypical of the full set of initial nonrespondents. Following Groves' (p.147) reasoning, four imaginary studies are presented in table 2.2. In study 1, with moderate efforts, a response rate of 50% has been achieved and the average duration of sports participation per week is calculated as 2 hours. In study 2, through extended field efforts active people who spend lots of time outdoors are contacted, and persuaded to participate, although they initially say they are too busy to answer survey questions. With a 60% response rate, the average duration of sports participation per week is 2 hours and 40 minutes, a substantial difference. Neither study, however, includes the 40% nonresponding uncooperative, socially isolated people. Had they been included, it would have been clear that the actual sports participation was similar to the estimate from the low response rate study and that increasing response rates by including an atypical part of the nonrespondents increased bias. These examples are very extreme, as the response rates per group are either 100% or 0%. Study 3 and 4 give slightly more realistic examples in which the bias in a 48% response rate case (3) is 15 minutes, and in a 60% response rate case (4) 20 minutes per week. Far from implying that nonresponse rates are not important, this example is presented to emphasize that a single-minded focus on response rates is not a fruitful approach.

Table 2.2 preludes the discussion on the possible detrimental effects of nonresponse bias on survey estimates in section 2.3. This section will also distinguish between nonresponse bias and the impact of nonresponse on the precision of survey outcomes due. Preceding this, section 2.2 will outline how nonresponse rates can be computed. Section 2.4 presents a short excursion into sampling. The purpose of this section is to firmly root this study in the probability-sampling framework and to show that response rates in probability and nonprobability samples cannot be compared. Section 2.5 presents an even shorter excursion into survey mode as a prelude to section 2.5, which identifies the main dimensions of nonresponse research, namely noncontact and noncooperation. The final section describes the different perspectives on nonresponse research that dis-

tinguish response enhancers, nonresponse measurers and nonresponse adjusters. The present study is situated somewhere in the middle.

Table 2.2 Response rates and bias (imaginary studies)

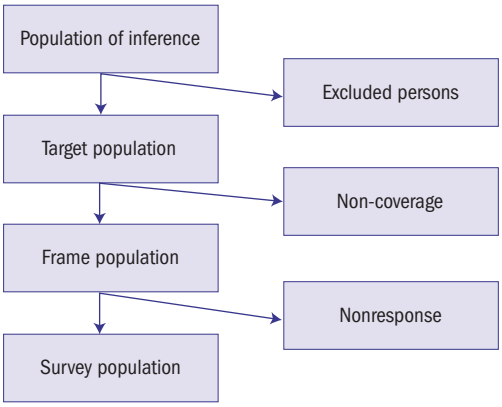
	share population	sports participation	study 1	study 2	study 3	study 4	full sample
	%	hours per week	response rates				
ordinary people	50	2	100	100	80	90	100
outdoors, active people	10	6	0	100	40	70	100
uncooperative, socially isolated people	40	1	0	0	10	20	100
total response rate			50	60	48	60	100
			estimate and bias				
estimate hours:minutes sports per week			2:00	2:40	2:15	2:20	2:00
bias in minutes per week			0	40	15	20	

2.2 *Computing and comparing response rates*

Groves (1989, chapter 3) starts his theoretical overview of populations with the *population of inference*, for instance Dutch citizens in 2000 (see figure 2.1). The *target population* is the finite set of persons that will be studied in a survey. From the target population those persons are generally excluded who cannot be contacted or will not be able to participate, such as persons living abroad and those living in institutions (homes for the elderly, prisons). The *frame population* is the set of persons for whom some enumeration can be made prior to the selection of the survey sample. After the sample has been drawn from the frame population, ineligible units have to be removed such as incorrect addresses or persons who are not Dutch citizens. Those who then respond to the survey are the *survey population*, the set of people who, if they had been selected to the survey, would be respondents. Unit nonresponse is the failure to collect data from units belonging to the frame population and selected to be in a sample. The response rate is the percentage of selected units who participate in the survey.

At every stage between *population of inference* and *survey population* groups drop out and operational and definitional problems have to be tackled. Excluding the homeless in advance means that they will not contribute to the noncoverage error when they are not in a sampling frame. Excluding the very elderly right from the start means that they will not contribute to the nonresponse error when they are not able to participate in a survey. A rich sample frame offering incomplete coverage may most likely result in

Figure 2.1 Types of population



higher response rates than a lean sample frame offering complete coverage (Groves and Couper, 1998, p. 299). Excluding difficult groups from the beginning may make life easier for those involved in fieldwork and will result in higher response rates, but will not make the survey any better. In assessing survey quality the total survey process should be taken into account.

Nonresponse in interview surveys may be due to no contact with the sample person (either because of physical impediments or not finding him or her at home), or because the sample person is not able or not willing to participate in the survey. A major problem in computing the response rate is that the sampling frame may contain errors and that it is not always easy to distinguish between ineligible sample units that are not part of the target population, noncontacts and noncooperation. In the Netherlands, as in this study, samples are often drawn from the Postal Addresses File (PAF). Groves (1989, p. 102) discusses definitional and operational problems of a list of housing units as the frame: 'Hours of discussion can be spent by those with years of experience in household surveys, trying to determine the appropriate status of particular living situations. In fact, however, most applications of the definition in the field are made by the interviewers or part-time staff form whom the listing operation is a small part of their full activities and who may have little experience in the task.'

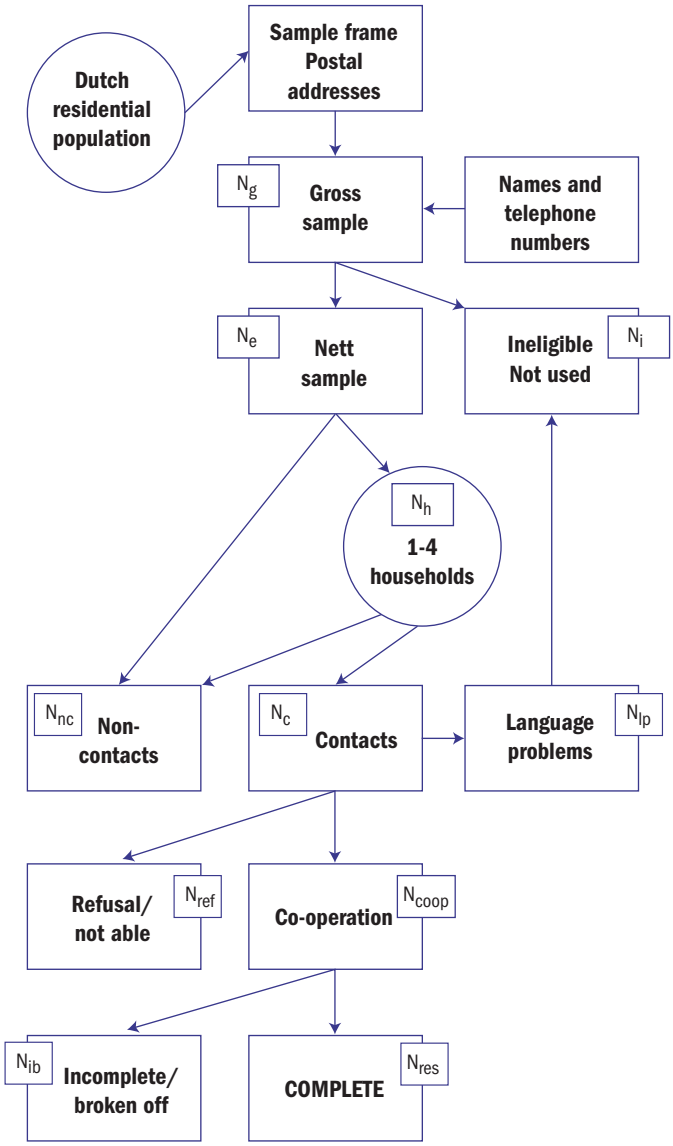
Response rates can be calculated in myriad ways (Groves, 1989, pp. 136, 140-145; Jowell, 1998, p. 171). Smith (1999, 2002) gives a both hilarious and depressing overview of outward attention to nonresponse standards, and the widespread disregard for or misleading information on response rate calculation and standards by government

agencies, academics, market research organizations and data archives. The American Association for Public Opinion Research (1998) provides standard outcome codes, distinguishing for instance between the respondent being dead, permanently physically unable to participate (senile, blind, deaf), and temporarily physically unable to participate (pneumonia, drunk). Lynn, Beerten, Laiho and Martin (2001, 2002) have developed standards to define final outcomes and response rates for the United Kingdom. One of their aims was that '... eventually commissioners of surveys will specify and assess response rates on the basis of these standards, thus improving the utility, validity and fairness of comparisons and judgements. This is particularly important when surveys are commissioned via competitive tendering' (2001, p. 2, web-document).

Figure 2.2 outlines the calculation of rates based on a Postal Addresses File (PAF) as sampling frame in a face-to-face survey. This scheme can be seen as an extension of Groves and Couper's (1989, p. 26) model of survey participation, and represents the sample and fieldwork procedure used in the AVO survey that figures prominently in later chapters. It starts with a random sample of addresses from which postal boxes and firms have been excluded, but addresses of shops have been left in because of the unlikely chance that households share their mailbox with a shop. This will increase the number of ineligible addresses but decrease non-coverage. At each eligible address the number of households has to be assessed, which is generally one. The row of doorbells at a multi-unit address may indicate the number of households, but in other cases contact with and cooperation from at least one resident is required. In this particular case, up to four households per address had to be interviewed. After contact has been made, language skills have to be determined, as non-Dutch speakers were considered ineligible (see Groves, 1989, p. 137-138). In the AVO any responsible adult was eligible for answering a household questionnaire, and subsequently all household members of 6 years and older had to fill in a drop-off self-completion questionnaires. Only if all drop-off questionnaires were completed, a sample unit was considered as response. Figure 2.2 shows how the contact rate, the cooperation rate and the response rate can be calculated according to this scheme. It does not show how difficult decisions are made in the field on how to classify different outcomes.

The use of standard outcome codes is far from common procedure in survey research, and using these codes in the field is far from simple. Especially the eligibility of selected sample units appears to be problematic, as it may be hard to separate from noncontacts. The American Association for Public Opinion Research (AAPOR, 2000) has reserved a special category (3.20) for cases of unknown eligibility and no interview

Figure 2.2 From population and address sample frame to household response



N_g ~ gross sample of addresses
= eligible (N_e) and ineligible (N_i) addresses

N_h ~ number of households in sample (1-4 per address)
= noncontacts (N_{nc}) + contacts (N_c)

N_c ~ number of contacted households
= number of refusals/not able (N_{ref}) + initial cooperation (N_{coop})

N_{coop} ~ number of initially cooperative households
= incomplete households, broken-off (N_{ib}) + complete sets (N_{res})

Cooperation rate = N_{res}/N_c

Contact rate = N_c/N_h

Response rate = N_{res}/N_h

in which a housing unit exists, but it is unknown whether an eligible respondent is present in the household unit.[1] This category is of course not very helpful in computing a response rate. Groves (1989, p. 139) describes the determination of eligibility in face-to-face surveys: 'In practice, the disposition of each sample number is determined by survey field personnel. The application of the classification system is, however, subject to error. These errors arise because of both the inherent difficulty in determining the appropriate outcome and influences of evaluation procedures to misclassify results.' He gives a number of examples that will sound familiar to everybody who has ever discussed how to compute response rates.[2] Biemer et al. (2001) present the results of an analysis of Post Enumeration Surveys to determine overcoverage and undercoverage in the US Census. Their study illustrates the difficulty of assessing the residence status of household members and, as a consequence, the possibility of misclassification if residency would be used as a determinant of eligibility.

A common practice in the Netherlands is to subtract unprocessed cases from the gross sample. An analysis by Bethlehem and Schouten (2003) of nonresponse in a Statistics Netherlands survey showed that in 1998 6.4 % of the sample units had not been processed due to lack of capacity (too high a workload for the interviewer) and interviewer unavailability (illness, holiday). The highest percentages of unprocessed cases occurred in interviewer districts in large cities. From these results it seems likely that excluding sample cases from the gross sample, and thus equating them with ineligible cases, will cause bias if the underlying lack of capacity is related to respondent characteristics and possibly to survey variables. Schnell (1997, pp. 23-24)

1 The federal Committee on Statistical Methodology (2001, p. 4-3) phrased this problem in this way: 'One difficulty inherent in response rate calculations is that the eligibility of a nonresponding case may sometimes be difficult to determine. For example, it can be very difficult to discern whether telephone numbers that are never answered are nonworking numbers versus households with no one at home. Similarly, mail surveys not returned by the postmaster are assumed to have reached the correct household or business establishment, but they may not have been received.'

2 'Many of the results which identify cases that are not part of the target population can easily be confused with nonresponse among eligible sample cases. For example, many houses that are vacant for the entire survey period are indeed retained for use by an eligible household. Families take extended vacations; temporary job assignments force people to leave their principal residences for short periods. An interviewer can repeatedly call without contact on a housing unit with a broken doorbell, residents who are hard of hearing, or occupants afraid to open the door to a stranger. After repeated unsuccessful efforts, interviewers might feel justified in coding the case as a vacant unit. If, in addition, there are no clear signs of habitation, the pressures to avoid a nonresponse might be overwhelming. Since individual interviewers are evaluated on the ratio of interviews to total eligible sample units they are assigned, they have an incentive to reduce the denominator of the ratio. Miscoding noncontacts cases as vacant units produces this result.' (Groves, 1989, 139-140).

mentions similar problems in Germany. Therefore, unprocessed cases should be considered as administrative nonresponse rather than ineligible cases.

The temptation to label sample units that do not provide an interview as ineligible is very strong. To define persons above a certain age limit, nonrespondents in a parent survey, dwellings without obvious inhabitants, non-Dutch speakers and unprocessed sample addresses as ineligible means that the response rate will be higher, and the higher the response rate presumably the better the survey and the happier survey funders, survey researchers and journal editors will be. As mentioned above, however, increasing response rates by removing problematic groups from the denominator does not improve survey quality.

How response rates are computed is particularly relevant when comparing response rates across survey modes or across countries or over time. Next to the size of the response rate, many researchers are concerned about a decline in response rates. More than 10 years ago Bradburn said in his Presidential Address at the 1992 meeting of the American Association for Public Opinion Research: 'We all believe strongly that response rates are declining and have been declining for some time'. Empirical evidence for declining response rates is not unambiguous, however. Smith (1995, p. 158) contests 'But despite the overwhelming consensus that response rates have been falling and will continue that decline into the future, the empirical evidence from the U.S. is both more equivocal and less uniform.' Empirical evidence may not be clear-cut for three major reasons (see also Groves and Couper, 1998, p. 158-159, who present several reasons why comparing response rates over time is difficult). Firstly, nonresponse rates are not always computed, computed uniformly or publicized, as Smith (1995) has shown convincingly (see also Schnell (1997, 48-56)). A second reason is that sampling and fieldwork procedures of a survey may change considerably over time. Furthermore, nonresponse rates may be similar, but the composition of nonresponse may change (Groves, 1989, p. 146). And finally, it may well be that response rates have been maintained at a similar level due to increasing fieldwork costs. As Bradburn said later in his Presidential Address: 'The problem of declining response rates is better described as a problem of increasing costs to get the same response rates.'

De Heer (1999) compared response rates across countries and found that despite substantial differences in survey design, the Netherlands scored uniformly low on each aspect of response. He partly confirms the decline in survey response, although trends differ from country to country, and strongly advocates setting up a databank providing information on response rates, types of nonresponse and survey characteristics to really compare response rates between surveys, cross-nationally and over time. De Leeuw and De Heer (2002, p. 52) studied response trends for a comparable series of surveys and concluded that '(1) countries differ in response rates, (2) the response rates have been declining over the years, and (3) the nonresponse trends

differ by country'. They ascribe the majority of these cross-national differences to differences in survey design and fieldwork strategies, especially the supervision and monitoring of interviewers. Taking all evidence and measurement issues into account, Groves and Couper (1998) conclude that longitudinal response trends are not as clear as many seem to think (p. 164) but that cooperation with surveys is declining in the US (p.166). Steeh et al. (2001, p. 242) conclude that nonresponse is not getting much worse, but is also not getting better. They do find a change in the composition of nonresponse in metropolitan areas, where refusals declined and noncontacts increased. Schnell (1997) tried with great efforts to dig up response rates on surveys in Germany from 1970 to 1990. He found a decline of 5% over 20 years, which was much less than expected (p. 130). Curiously, he did not find any evidence for increased field efforts over time, except for an increase of the number of contact attempts. Mohler, Koch and Gabler (2003) wonder why nowadays a response rate of 50% seems the maximum attainable in Germany. They call for a discussion on fieldwork procedures in Germany as they consider the current situation to be not just a case of bad luck but a case of deficient fieldwork organization and insufficient interviewer remuneration. Investing in the supervision and monitoring of interviewers might make it possible to break out of the downward trend and get back to an acceptable level.

In the Netherlands, as in the US, evidence on response rate is neither unequivocal nor uniform. Response rates on the National Election Study went down from close to 70% in the seventies to below 50% in the 1990s, leading to heated discussions on the relationship between political interest and response rates (see section 4.1.1). Response rates on surveys run by Statistics Netherlands showed a decline but seem to have recovered recently, most likely due to closer fieldwork monitoring. The AVO-survey of the SCP is back at its original level after a serious dip in the early 1990s, due to increased field efforts. The European Social Survey (see chapter 10), shows that response rates close to or even exceeding 70% are possible in many European countries, even in the Netherlands, a level that many would not have considered to be feasible anymore.

2.3 *When and why nonresponse can be a problem*

Bethlehem (2002) presents a model for nonresponse, outlines when nonresponse can cause bias and when nonresponse can be adjusted by weighting with auxiliary information. In his formal notation he starts with a random sample of size n that is selected from a population. Let y_1, y_2, ... y_n denote the values of the target variable for the n selected elements. Furthermore, it is assumed that each unit *k* in the population has a certain, unknown probability ρ_k of response.

Let

$$\bar{y}^* = \frac{1}{m}\sum_{i=1}^{m} y_i$$

denote the mean of the m (m<n) available observations. This, however, is not an unbiased estimator of the population mean, as the expected value is approximately equal to

$$E(\bar{y}^*) \approx \frac{1}{N}\sum_{k=1}^{N} \frac{\rho_k Y_k}{\bar{\rho}}$$

Therefore, the bias of this estimator is approximately equal to

$$B(\bar{y}^*) \equiv E(\bar{y}^*) - \bar{Y} \approx \bar{Y}^* - \bar{Y} = \frac{C(\rho, y)}{\bar{\rho}}$$

in which

$$C(\rho, y) = \frac{1}{N}\sum_{k=1}^{N} (\rho_k - \bar{\rho})(Y_k - \bar{Y})$$

can be seen as the population covariance between response probabilities and the values of the target variable. From the expression for the bias is follows that the estimator is approximately unbiased if there is no correlation between the target variable and response behaviour, and the stronger the relationship between target variable and response behaviour, the larger the bias. Furthermore, the size of the bias depends on the amount of nonresponse and the contrast between respondents and nonrespondents.

To correct for a nonresponse bias, adjustment weighting is often carried out. Each unit i is assigned an adjustment weight w_i which leads to a new estimator of the mean:

$$\bar{y}_w = \frac{1}{n}\sum_{i=1}^{n} w_i y_i$$

Weighting adjustments are based on auxiliary information, i.e. variables that have been measured in the survey but for which information on the sample distribution or the population distribution is also available. This could be sample frame data, interviewer observations, information from low-level geographic databases or population distributions. Weighting based on auxiliary information will improve estimates for target variables if there is a sufficiently strong relationship between auxiliary variables and target variables. If there is no such relationship weighting will not improve the estimates. Elliot (1991, p. 18) even gives examples where weighting will result in worse estimates.

A related way of looking at the possible impact of nonresponse is by distinguishing between *precision* and *accuracy*, two concepts within the framework of total nonresponse error (e.g., see Groves (1989) in the first chapter of his book on survey errors and survey costs). The total nonresponse error is the difference between the outcomes of a survey with and without nonresponse. The Mean Square Error (MSE) of the actual outcome of the survey (with nonresponse) is the expected difference between all possible realizations of the survey and the true value (without nonresponse). The higher the MSE, the smaller the accuracy of the outcomes. The MSE comprises two components, the *variance* and the squared *bias*. The *variance* indicates the precision of the outcome. If nonresponse in a survey is high but is not related to the topic of the survey, the outcomes of the survey across all possible realizations will show a large variation, and the larger the variance, the lower the precision. If nonresponse affects the precision of a survey, this can be mended by increasing the sample size. The *bias* is the type of error that affects the outcome in all implementations of the survey design. If there is nonresponse bias, i.e. there is no random process behind response, increasing the sample size does not help. This is equally true for increasing the sample size for specific groups that are known to have low response rates only (see Lynn (2004) and the discussion on booster samples in section 2.4.2).

A third way of assessing the impact of nonresponse and distinguishing between random effects and bias are the three types of missing data from Little and Rubin (1987).[3] Firstly, units may be missing completely at random (MCAR). This would happen if bad sectors of a floppy disk make a random part of the data inaccessible, or if interviewers who have their birthday in the fieldwork week send their assignment back unused. MCAR nonresponse would reduce precision due to smaller final sample sizes. Precision can be increased by increasing the initial sample size. However, this simple situation is extremely unlikely. A more realistic assumption is that units are missing at random (MAR). This would for instance occur if women responded less often than men, but there were no difference between responding and nonresponding women. A larger sample size could maintain precision. Oversampling of women or weighting by sex can adjust the estimates.

The worst case, non-ignorable nonresponse, occurs when survey variables are related to response propensity and are not missing at random (NMAR). In this case nonresponse is related to survey variables that are not measured for nonrespondents,

3 Little and Rubin discuss three types of item nonresponse. This occurs when questions that should have been answered are not. Their distinction is equally valid for unit nonresponse, the object of this study. Item nonresponse will only be discussed as an aspect of survey quality in section 7.4. For that matter, the distinction between item nonresponse and unit nonresponse is not clear-cut. A questionnaire containing to much item missings could be rejected, for instance, and be considered as unit nonresponse.

for instance when those interested in politics are eager to participate in election surveys, socially isolated persons refrain from answering questionnaires on social participation, crime victims do not open the door to an interviewer, hospital patients cannot participate in a health survey, and healthy, out-door people cannot be contacted with the request to participate in a time use study. In this case, nonresponse will not only decrease precision, but also increase nonresponse bias (Lynn et al., 2000, p.135). In the NMAR case, increasing the sample size or weighting for nonresponse does not solve the problem and thus does not produce more accurate estimates. This implies that nonresponse rates are not necessarily the main problem, but bias due to nonresponse. This will be the underlying theme of this study.

Summarizing, three types of nonresponse can be distinguished, MCAR, MAR and NMAR. Firstly, nonresponse can be at random, in which case it will reduce precision. Simple remedies are trying harder by extending the fieldwork period or making more calls, both of which will increase fieldwork costs. One problem is that it will not always be clear in advance when nonresponse is just a random factor. One might assume, for instance, that random, situational factors (see section 4.1.2) such as having a cold or the milk boiling over when the interviewer calls can happen to everybody, and may lead to a loss of precision rather than bias. This would be a case of missing at random (MAR), or perhaps even completely at random (MCAR). In a health survey, however, the nonresponse of those with a cold might give a biased estimate of health, and in a survey on household coping behaviour, the nonresponse of persons with limited cooking skills (causing them either not to cook or have the milk boiling over) might give a biased estimate of coping behaviour, and these would be cases of not missing at random (NMAR).

In the second type of nonresponse (MAR), specific identifiable groups in society – possibly dependent on survey mode and survey topic – might have a lower probability of cooperating in a survey. These groups could be identified by socio-demographic, socio-economic or other characteristics on which independent information is available from the sample frame or from population distributions. In this case weighting for nonresponse can reduce bias, if there is a strong relationship between target variables and these auxiliary socio-structural variables, and if there is not a strong relationship between target variables and response behaviour independent of these socio-structural variables. The statistical price that may have to be paid for weighting is an increase in the variance of estimates, and thus precision (Lynn, 1996, p. 211). An alternative remedy, which would be more costly, is to oversample among these groups that have higher nonresponse probabilities, for instance the inhabitants of big cities.

Neither solution, weighting nor oversampling, will work for the worst case of nonresponse when response behaviour is directly related to target variables of the survey (NMAR). The only way to find out if there is non-ignorable nonresponse and – pos-

sibly – to correct for it, is to obtain as much information as possible on nonrespondents. Bethlehem (2002, p. 287), for instance, suggests, using sample frame data, interviewer observations, statistical agency records and, as a second source, obtaining information from the nonrespondents themselves by the method of Hansen and Hurvitz (1946) and the Basic Question Approach of Kersten and Bethlehem (1984). These approaches will extensively be discussed in chapter 5.

2.4 *Nonresponse and sampling*

2.4.1 *The perfect survey: a nightmare*

Daydreaming about the perfect survey, one comes up with a long questionnaire in which a wide range of issues is tackled in-depth with well-tested theory-driven, reliable and valid questions, to be administered to a very large sample of the population to allow a distinction to be made between social and regional subgroups, including the institutional and non-institutional population, the homeless, illegal aliens and speakers of minority languages, to which all selected sample persons respond graciously, and the results of which are freely available for everyone immediately after data collection. This ideal survey might rapidly transform into a nightmare, but is fortunately inherently impossible. It would be prohibitively expensive, and data collection according to these extreme standards would rule out timely availability of the data. The homeless, the very elderly in senior housing facilities, inmates of prisons and illegal aliens might have trouble participating for diverse reasons, might not consider the questionnaire very relevant for their situation and might necessitate different fieldwork modes which make subgroups incomparable. Questions delving into the intricacies of Dutch society might be difficult to translate into minority languages. And finally, there will always be refusal to cooperate, except in a mandatory survey run by a dictator regime where the veracity of the answers might suffer from the penalty on noncooperation.

Daydreaming about high response rates, one could envisage drawing a sample from a group of cooperative, easy to reach individuals who have agreed to fill in surveys regularly. One could also envisage a very short questionnaires on interesting subjects to be administered to respondents in their own preferred survey mode, with no limits on the duration of fieldwork and lavish incentives to hand out to respondents and interviewers. The first option would be tantamount to striving for a high participation rate among highly participative individuals, which might not be a major feat. The second option would probably result in admirably high response rates, but also reduce the content of the survey, endanger the timeliness, increase coverage error and make it very expensive. Very high response rates are possible but it would be wrong to focus on enhancing response rates alone. A survey with very low nonresponse rates, based on a non-probability sample in which many groups are not covered, using a very short, pleasant but non-informative questionnaire is not a good survey just because response rates are high. Voogt (2004, p. 117) showed that an astonish-

ingly high response rate of 93% is possible, if one is willing to reduce the questionnaire to two questions for the most difficult respondents. This provides very useful information for adjusting for nonresponse, but two questions may not always provide enough content for substantive research.

What remains from these daydreams are a number of survey quality criteria that jointly determine the quality of a survey, and the realization that there always will be a trade-off between different criteria. The first criterion refers to the accuracy of the outcomes. This is dependent on four types of survey errors (Groves, 1989):

- *coverage errors*, which occur because each member of the population concerned is not given a known non-zero probability of being included in the survey sample;
- *sampling errors*, which occur because only a subset of the population is surveyed;
- *measurement errors*; which occur because the measurement vehicles (such as the questionnaires, interviewers or coders) introduce faulty answers
- *nonresponse errors*, which result from respondents and nonrespondents having different characteristics that impinge on the survey's purpose.

Furthermore, accessibility, interpretability, coherence and timeliness are important survey quality criteria. The main quality criterion, however, is the relevance or content of a survey. Accurate data on trivial issues do not have much value. In recent overviews of survey quality criteria (Lyberg, 2001; Lyberg et al., 2001; Fellegi, 2001 and ONS, 2004) the importance of a survey meeting users' needs is emphasized. These should be established at the outset, which will be easier for a survey aimed at a single user than for a survey catering for divergent needs of different types of users. Different users might focus on national relevance, or cross-national comparability, or longitudinal continuity. Some users will concentrate on specific issues and need questions that go into detail, other may set a higher value on a broader coverage of more issues, or a larger sample comprising distinct subgroups. What users consider relevant may not be what respondents consider relevant, or salient (Morton-Williams, 1993, pp. 26-27; Goyder, 1987, pp. 118-130), so policy issues that are highly relevant for the researcher may nonetheless put off potential respondents and result in low response rates. In European statistical organizations the criteria mentioned here have generally been adopted as quality criteria of official statistics.

This study focuses on nonresponse and nonresponse errors and generally ignores other survey quality criteria. In the subsequent sections a short side-step will be made into the area of sampling. The main reason for this is that online access panels, generally based on nonprobability sampling, are nowadays often presented as the solution for the nonresponse problem.

2.4.2 *Representative samples and probability samples*

In the nonresponse literature a sample is generally assumed to be a *random or probability sample*. This has not always been the case. In her overview of the history of survey research in the United States, Converse (1987) shows that in the earlier surveys (random) samples were not much of an issue. She outlines the permanent controversy between market research organizations, where *non-probability samples* such as quota sampling (see below) and 'juries' are routine, and the scientific surveys from official statistics, where the theories from inferential statistics lead to no other choice than probability samples. Fowler (2002, pp. 53-56) points to the same divergent opinions on acceptable sampling methods: 'The federal government will not fund survey research efforts designed to make estimates of population characteristics that are not based on probability sampling techniques. Most academic survey organizations and many nonprofit research organizations have a similar approach to sampling. At the same time, most of the major public opinion groups, political polling groups, and market research organizations rely heavily on nonprobability sampling methods' (p.53).

This controversy goes back to the conflict between purposive selection and random selection in the first half of the 20th century, described by Kish (2003, pp. 8-9). In May 1924 the International Statistical Institute (ISI) appointed a commission for the purpose of studying the application of the *representative* method in statistics. This commission distinguished in its report random selection (meaning here identical inclusion probabilities for each unit) and purposive selection, meaning that a number of groups of units are selected that together yield nearly the same characteristic as the totality. The concept of 'representativeness' is central. Kruskal and Mosteller (1979a, 1979b and 1979c) present a complete and diverting overview of what representative is supposed to mean in non-scientific literature, scientific literature excluding statistics and in the current statistical literature, and give the following meanings for 'representative sampling': 1) general acclaim for data, 2) absence of selective forces, 3) miniature of the population, 4) typical or ideal case(s), 5) coverage of the population, 6) a vague term, to be made precise, 7) representative sampling as a specific sampling method, or 8) as permitting good estimation, or 9) as good enough for a particular purpose.[4]

4 Readers will be reminded of Borges' classification of Chinese animals apparently based upon a certain Chinese encyclopaedia. Here animals are divided into those that belong to the Emperor, embalmed ones, those that are trained, suckling pigs, mermaids, fabulous ones, stray dogs, those included in the present classification, those that tremble as if they were mad, innumerable ones, those drawn with a very fine camelhair brush, others, those that have just broken a flower vase, those that from a long way off look like flies.

An example of making a sample representative according to meaning 4 is taken from an article in the New York times describing the selection of a White House Conference on Youth in 1971 (cited in the 1979a paper, p. 20): '...The 100 youths on each panel are supposed to fit into sexual, ethnic, vocational and regional patterns. Fifty should be male, 50 female. There should be 16 college students; 39 students at vocational, trade or high schools; 39 individuals who are not in school, and 6 in the armed forces. Each panel should also contain 70 whites, 6 Mexican-Americans, 3 Puerto Ricans, 3 'other Spanish', 14 Negroes, 2 American Indians and 2 Oriental-Americans. In addition each panel; should reflect the population percentages in each region of the country.' 'The biggest problem now is we have to find poor, white working females' said one staff official.' Kruskal and Mosteller wonder how cross-classifications should be handled, given these distributions (how many Puerto Ricans should be male?). In their 1979b paper the authors quote W. Lewis Hyde giving an example of the 3rd meaning of representative (p. 120): 'The Chairman said, 'Are you sure that the parking lot contains a truly random sample of modern society? 'Well, maybe not,' said the O.E., 'but we did the best we could. We generated a selection list using a table of random numbers and a set of automobile ownership probabilities as a surrogate for socio-economic class. Then we introduced five racial categories, and an equal male-female split. We get a stochastic sample that way, with a kind of 'Roman cube' experimental protocol in a three-parameter space.' 'It sounds complicated,' said the Chairman. 'Oh, no. The only real trouble we've had was when we had to find an Amerindian woman driving a Cadillac...'

It is tempting, but perhaps not altogether acceptable, to cite all three papers by Kruskal and Mosteller including the part where they make recommendations about scientific usage for each meaning (p. 125, it boils down to 'don't use the word *representative*, but specify what you mean'). One would expect that after these papers nobody would dare to use the word representative again with regard to sampling and indeed Kish (2003, p. 12) felt in 1995 that *representative sampling* was a term that can be avoided and assumed it was disappearing from the technical vocabulary. Schnell (1997, p. 12) calls it an immeasurable, unscientific concept, but in the Netherlands it is still a household word in describing samples and sampling although it is not clear whether this means a miniature of the population, or good coverage of the population, or a random sample or something else.

In probability sampling it is necessary and sufficient that every member of the population has a known probability greater than zero of being included in the sample. When the inclusion probabilities are not equal, design weights have to be used (Kish, 1994, pp. 161-164). The literature on probability sampling is extensive.[5] Survey estimates from a random sample are unbiased and their precision increases with the sample size. From probability samples inference to the underlying popula-

5 Useful introductions in probability sampling are given by Kalton (1983) and Bethlehem (2000) (in Dutch).

tion is possible. Probability samples can be selected for a single survey once, for a survey and a follow-up, or for a series of surveys. *Single-phase samples* are, for instance:

- simple random sampling, with equal inclusion probabilities, where each household (or more correctly each address) has an equal probability of being included;
- two-stage random sampling (e.g. first communities, then households);
- stratified random sampling;
- booster sampling (oversampling in areas where response rates are expected to be low);
- cluster sampling, in which all members of a sampling unit are interviewed, for instance all pupils in selected school classes.

In simple random sampling, the probability of every member of the population being included is identical. In the other types, the inclusion probability will differ. For instance, in stratified random sampling, samples are drawn within strata without the sample size reflecting the population sizes of the strata. An example is a European survey in which the sample size is identical in every country. Here the inclusion probability of Luxembourgers and Slovenians is much higher than of the Germans and the French. Stratified random sampling is sometimes confused with quota sampling and sample strata are sometimes erroneously designated as quota (Kalton, 1987, p. 92).[6] Another type of stratified samples are booster samples (oversampling), where the sample size in areas or among groups with an expected low response rate is boosted (see for instance Lynn, 2004). The larger initial sample should compensate for the lower response rate. This procedure implies that one resigns oneself to lower response rates among specific groups (for instance big city inhabitants) in advance. It improves precision, but might be risky when it only succeeds in increasing the number of, for instance, middle class big city inhabitants in the sample rather than obtaining the participation of lower-class inhabitants of big cities who caused the low response rates to begin with.

Probability samples can also be interviewed in *two phases*. Examples are a face-to-face interview with a drop-off self-completion questionnaire for all respondents, or a telephone survey succeeded by a face-to-face interview of a specific subset of respondents, or a 'screener survey' conducted solely to identify specific groups that are subsequently interviewed. If the first survey, based on probability sampling and often called a parent survey, is used to select a specific group for further questioning, the two-phase survey is similar to routing within a questionnaire. If respondents in the parent survey do not cooperate in the second survey, this might be considered item

6 In quota sampling any female, young, working inhabitant of Amsterdam will do to fill the quotum of female, young, working inhabitants of Amsterdam. In stratified sampling, female, young, working inhabitants of Amsterdam can form one of the strata (provided a rich sampling frame is available). Within this stratum a random sample is drawn of which – ideally – every person should cooperate. Here, only those individuals that are part of the selected sample can be interviewed, and the interviewer does not stop when a fixed target has been reached.

missing. In the second survey, initial nonresponse should be taken into account. Randomly selected sample members can of course also be interviewed several times. This happens in *longitudinal panels*. The selected sample persons have to answer several questionnaires over time instead of one.

2.4.3 *Access panels as a solution to the nonresponse problem?*

There are many forms of non-probability sampling (see for instance Kalton, 1983; Couper, 2000; Schonlau, Fricker and Elliot, 2001). Kalton (1983) distinguishes convenience samples (generally comprising volunteer respondents), purposive samples (the researcher picks units that are 'representative', similar to the original idea of Kiaer) and quota samples, which are very much *de rigueur* in France, even so that French researchers can be sincerely surprised that the response rates in a face-to-face survey are not 100 per cent. Nonresponse is not relevant in quota sampling [7], as the model specifying the quota controls excludes a direct relationship between response propensities and survey variables and as all individuals who conform to the selection criterion for a quota are considered to be interchangeable.

Access panels are groups of people who have agreed to regularly participate in surveys run by a specific organizations, generally a market research organization, and are the modern equivalent of mail panels. They are becoming increasingly popular in the Netherlands and are advertised as instruments for drawing high response samples from a very rich sample frame, comprising information from previous surveys among the panel members.[8] Access panels can be based on probability and on non-probability sampling (Schonlau, Fricker and Elliot, 2001; Couper, 2000). The former are also called pre-recruitment panels, the latter volunteer panels or convenience samples. In pre-recruitment panels recruitment is passive. After completing a survey, the respondent is invited to become an access panel member, and after each survey the panel gets larger. If the recruitment is from random sample surveys, and data on initial inclusion probabilities and response rates are available (which is rarely the case), a survey among (a random sample from) an access panel can theoretically be considered as based on probability sampling. If inclusion probabilities and nonresponse in earlier phases are unknown, and if there is a large attrition in earlier phases, the probability character of access panels is easily lost. In this case it is not clear any more what the target population is and one cannot maintain that the inclusion probability of every member of the target population is known and greater than zero. In a volunteer panel recruitment is active: respondents can apply for membership of an access panel. When there are tight controls on panel membership, a

7 Deville (1991, p. 175) developed a theoretical way of estimating inclusion probabilities: 'Thus, estimation techniques based on fitting should allow for the honourable processing of non-responses in quota surveys.'

8 See for instance http://english.tns-nipo.com/ and www.gfk.nl/data/instrumt/instrumt.htm consulted on 9 August, 2004.

volunteer panel can be quite similar to a quota sample. Both pre-recruitment and volunteer panel will be treated here as forms of non-probability samples here under the heading of access panels.

Access panels have several important advantages, earlier described by Stoop (2004b).[9] Firstly, they are often recommended for their high response rates. Hox and De Leeuw (1994, p. 337) confirm from a multilevel study in which they compared response rates across survey characteristics that '... Samples that are national and random receive a lower response rate than local and nonrandom samples...' Nonrandom samples in their study are convenience samples or existing panels. This does not seem to a fair comparison, however, as in convenience samples and access panels initial nonresponse is ignored, and access panel members have stated in advance that they are willing to participate in future surveys. A second advantage is that samples can be drawn from access panels that are similar to the population in terms of background characteristics and that will remain similar because of these low nonresponse rates. Weighting for underrepresented groups (because of sampling errors, noncoverage and nonresponse) will no longer be necessary . This is related to the third advantage, namely that specific subpopulations can easily be selected from access panels as many characteristics of the panel members are known, or because screening is relatively inexpensive. As a consequence, which would be advantage four, substantive information from panel members may be known on characteristics that are generally considered to be related to survey nonresponse and that can be used in drawing a sample, such as political interest, voting behaviour or social isolation. A fifth advantage is that the topic of the survey is of smaller importance in determining survey cooperation than in ad hoc samples. Access panel members are generally favourable towards survey research and there decision to participate will be less dependent on their interest of one survey topic or their dislike or lack of interest for another. The sixth advantage is that the turnaround time of access panels is generally shorter than for random samples. In the latter case, many calls have to be made to contact a sample household (in face-to-face surveys), mail or e-mail reminders have to be sent, and initial refusers have to be converted.

In access panels, especially when they are Internet panels, the response process can be speeded up substantially. Panel members are expecting their questionnaires and are interested and willing to participate from the start. The final advantage is related to the previous ones: surveys using access panels are often less expensive than random samples. This may be due partly to the fact that access panels generally use web-based or mail surveys. They are presumably also less expensive, because panel

9 There may be other advantages and disadvantages besides the ones mentioned here that are a consequence of the interview mode, as many panels use Computer Assisted Self Interviewing (CASI) via the Internet. The interview mode is independent of the non-probability characteristic of the sampling in an access panel, however.

members have promised to regularly take part in surveys, and therefore less time has to be devoted to persuading respondents to participate. In general, incentives are given to access panel members as a sign of appreciation for participating and to keep them in the panel.

Access panels also have serious limitations. Firstly, access panels may attract professional respondents who are in it for the rewards. Pring (2005) states that '... members of the panels of each of the 10 leading survey panel companies (in the US) belong on average to 7 other survey panels'. Secondly, statistical inference to the underlying population is not possible with nonprobability samples. If there is no strictly random mechanism through which individuals can become part of an access panel, and if inclusion probabilities are unknown, results from an access panel refer to the access panel and cannot be generalized to the population.

The *third* reservation regards the claim of high response rates. This ignores initial nonresponse, which may be quite high. Sudman and Kalton (1986) discuss the use of mail panels as a means of sampling rare populations. They conclude: 'Although 80% to 90% of panel households cooperate on a study, the major problem with mail panels is that the initial cooperation rate of households invited to participate in a panel is often 10% or less. Mail panels are usually balanced by major demographic variables to remove the most obvious selection biases, but other biases still remain. These unknown selection biases may distort the survey results, and the researcher will not be able to assess the possible distortion unless some independent check can be made'. Defining response rate as the response propensity of willing respondents makes response rates difficult to compare with those of probability samples (or stated otherwise: response rates among those who are willing to participate in probability samples will be very high too). As was shown in section 2.3 high response rates are a pursuable aim but high response rates among willing, easy to contact access panel members do not solve the problem of nonresponse bias. In access panels, high response rates may simply hide the problem that less willing, participative groups are left out. This effect may be mitigated by the fact that the topic of the survey may be less crucial for survey participation than in an ad hoc survey, and by the availability of substantive variables that are generally related to nonresponse, but if and to what extent this is the case will remain unknown, especially as there are multiple reason for nonresponse.

Therefore, access panels 'represent' the population if socio-structural or background variables are related to survey variables, and there is no relationship between panel membership and survey variables independent of background variables. An illustrative example of the perceived importance of background variables can be found on the website of TNS-NIPO, which maintains an access panel of 200,000 individuals that can be approached through diverse survey modes and says about its Computer Assisted Self Interviewing panel: 'The names and addresses of these pc-owners have

been collected over the years by TNS NIPO and screened using a variety of research instruments. Extra efforts have been made to create a considerable volume in the category of respondents with low pc-ownership figures (particularly older people and the less well educated). This makes it possible to take samples from the total database of, for instance, 5,000 people who are representative with regard to age, education, gender, et cetera' (http://english.tns-nipo.com/sub.asp?c026, 9 August 2004).[10] The underlying assumption here is that target variables depend on age, education, gender, etc. and that there is no strong relationship between both the likelihood of being a panel member and owning a PC and target variables. These assumptions are similar to the claims sometimes made with respect to a more traditional form of nonprobability samples, namely quota samples, that they are as good as or better than random sampling, because they can be made 'representative' or similar to the population in composition in advance. Morton-Williams (1993, pp.31-35) has shown that the latter claim is based on two assumptions, namely '... that the behaviour and attitudes to be measured are related primarily to the variables used as quota controls; secondly, that they are not associated independently of these controls with factors underlying nonresponse nor with the characteristics of those likely to require more than one call to obtain an interview' (p.32). Or, as Sudman (1966) said: 'In probability sampling with quotas (sic) the basic assumption made is that it is possible to divide the respondents into strata in which the probability of being available for interviewing is known and is the same for all individuals within the stratum, although varying between strata.' These assumptions are similar to those underlying the MAR type of nonresponse outlined above. They cannot be put to the test, however.[11]

What can be done is to compare the survey outcomes of an access panel and a random sample. These comparisons have been made by market research organizations but the outcomes are not always publicly available. Peters (2001) compared the results of the Cultural Changes in the Netherlands survey conducted among 500 panel members who filled in the questionnaire on their own PC and 1500 randomly selected face-to-face respondents. She found substantial differences. In this study it was not clear whether the differences were due to the interview mode (CASI or CAPI) or to the sampling procedure. Bronner, Tchaoussoglou and Ross (2003) re-approached 500 panel members by telephone and asked them about readership of Dutch weeklies. The CASI approach in the panel resulted in lower readership of prestigious Dutch media, which confirms that self-administered questionnaires are less sensitive to

10 As section 3.3.7 will show, pc-ownership may be related to survey participation. This does not imply that pc-owners from groups where pc-ownership is rare, automatically 'represent' non-pc-owners in this group.

11 Access panels may be a more sophisticated instrument as outcomes of earlier surveys will be available in addition to general background variables. It still cannot be tested, however, that there is no relationship between panel membership and survey variables independent of these known characteristics.

social desirability, but sheds no light on the difference between probability and non-probability samples. Faas (2003) compared the results of members of an online access panel and Internet users selected from a random sample. He found that these groups differed with respect to socio-demographic variables, questions on Internet use and (after weighting) for political attitudes. A general problem with comparing results from random samples and access panels (or quota samples) is that even similar outcomes in one case are not a guarantee of similar outcomes in other panels or on other topics or at a later stage.

A frequent misunderstanding about web surveys and access panels is that large numbers make a sample better. Couper (2000, pp. 480-481) discusses a self-selected web survey: 'We received more than 50,000 responses – twice the minimum required for scientific validity –' .... while the survey did not yield a random sample and the selection probabilities are unknown, 'this *does not mean that the survey cannot yield representative social science data*' (Emphasis in the original). They claim that the selection probabilities can be 'estimated' by comparing the distributions on standard demographic variables to official government statistics and applying weighting. This assertion is based on the assumption that matching two 'samples' on a variety of demographic characteristics will ensure that they also match on the survey variables of interest.' Not surprisingly, despite the large number of respondents, they did not resemble the US population on a number of key indicators. Dillman and Bowker (2001) say about web surveys: 'Conductors of such surveys have in effect been seduced by the hope that large numbers, a traditional indicator of a high quality survey (because of low sampling error), will compensate in some undefined way for whatever coverage and nonresponse problems that might exist. Large numbers of volunteer respondents, by themselves, have no meaning. Ignoring the need to define survey populations, select probability samples, and obtain high response rates, together provide a major threat to the validity of web surveys.' Couper (2001, pp. 173, 184) also pointed to the misguided assumption that large samples necessarily mean more valid responses. Only in the case of probability samples does an increase of sample size to an increase of precision. In nonprobability samples, no inference to the underlying population is possible, and larger samples do not necessarily give better estimates than smaller samples.

The limitations of nonprobability sampling mentioned here do not mean that access panels should not be used. Several authors have tried to outline when probability sampling is preferable and when non-probability sampling is a useful tool. These arguments are also valid for the use of access panels. Sudman (1966) presents a very well conducted study using quota sampling with detailed quota controls. He concludes that '... To be more explicit, where survey results will receive very sophisticated analysis or when critical decisions will be based on them, it will be worthwhile to pay a substantial cost to achieve high standards of sampling, processing, and control. Thus, the Census Bureau rightly has very high standards on the Current Popula-

tion Surveys. On the other hand, many exploratory studies do not require such high standards since the analysis may be more limited and the questionnaire may itself be a major source of error. Here quota sampling would be justified.' Deville (1991) who is clearly in favour of quota samples, ends his theoretical overview with the conclusion that using a speculative model in a survey means methodological risk-taking, which may be entirely appropriate if the users who commissions this survey explicitly agrees with the details of the specification. 'Official statisticians, on the other hand, are responsible for data that can be used by the entire society; and that can be used, in particular, in the arbitration of disputes between various groups, parties, and social classes. The use of statistical models <in sampling>, particularly econometric models that describe the behaviours of economic agents, may turn out to be very dangerous, partial, or affected by a questionable or disputed economic theory. Official statistics should not tolerate any uncontrollable bias in its products. It should carry out sample surveys using probabilistic methods.' These authors discuss sophisticated forms of quota sampling, using a model-based approach, where in many cases quota sampling and access panels will be used in a looser, less model-based approach which on the one hand may make the impact of the model used less predominant, but also undermines the assumptions on which non-probabilistic sampling is based.

One sensible thing to do when using access panels is to follow Fowler's admonition (2002, p. 56) and be transparent: 'If a researcher decides to use a nonprobability sample, however, readers should be told how the sample was drawn, the fact that it likely is biased in the direction of availability and willingness to be interviewed, and that the normal assumptions for calculating sampling errors do not apply. Such warnings to readers are not common. In many cases, nonprobability samples are misrepresented seriously, and that constitutes a serious problem for the credibility of social science research.'

2.4.4 *A case for probability sampling*

Increasing fieldwork costs are likely to bring about an increase in the use of access panels. Marsh and Scarbrough (1990), however, looked at the effective cost of random and quota samples (random interviews were 50% more expensive) and cited Deming (1960, p. 31): 'cost has no measure without a measure of quality, and there is no way to appraise objectively the quality of a (quota) sample as there is with a probability sample'. High nonresponse rates may also cause survey sponsors to yield to the temptation of nonprobability sampling, especially as the costs of random samples with high response rates are increasingly difficult to raise. As Lamas (2003), the vice-president of a Spanish media research organization expressed the dilemma: 'Probability sampling is the only scientifically based system and other sampling procedures cannot be regarded as scientific. But, can we still consider a probability sample with a response rate of 20% as being scientific? Is this still preferable to a quota sample?' Or, put differently, if response rates are getting so low, we might as well use an access panel (although the response rate here might be very low too, if initial nonresponse is included).

Despite the difficulties in achieving or maintaining high or even acceptable response rates, it would be unwise to throw in the towel and give up probability sampling. Sampling theory shows how probability sampling enables inferences to be made to the population. Acceptable response rates (see chapter 10) are still possible. And finally, as this study will show, there is much to be gained if we aim for combating nonresponse bias rather than simply focusing on response rates. In 'A response to the nonresponse problem' Bradburn (1992) gives his view on quota sampling. 'When the predominant mode of data collection was face-to-face personal interviewing, one of the most vigorous debates in the survey world was the degree to which one could relax the requirements of probability sampling and substitute quotas for certain demographic characteristics without undermining the validity of the inferences that could be made from the data. As my colleague Martin Frankel is fond of pointing out, a well-designed quota sample which relaxes probability sampling at only the last stage of sampling is like a full probability sample with a 33 percent response rate, because that is about the percentage of households that are interviewed on the first call. Quota sampling works well for many purposes, and controlled experiments between sampling methods often produce little or no differences in data. The difference, of course, is that probability sampling rests on the mathematical theory of sampling, while quota sampling does so only approximately. Thus with quota sampling we lose the theoretical basis for drawing inferences from the data, and we never know when (or how often) our quotas will lead us astray.'

In summary, there are success stories in which nonprobability surveys perform extremely well. As Kish (1998) says, however '... These better methods (of quota sampling) tend to be used when the results can be checked against outside sources, such as election forecasts. This is clever and 'cost effective', but should warn us against the likelihood of much greater biases from the unchecked results of quota sampling.' Nonprobability samples in general, and access panels in particular, appear to be an attractive substitute for random samples, considering increasing survey costs and decreasing response rates. As access panels comprise persons who have expressed their willingness to regularly cooperate in surveys, one might conclude that they make the nonresponse problem obsolete. That is not the point of view on which this study is based, however. This study presents tools for monitoring response and estimating nonresponse bias in random sample surveys. In view of the limitations of nonprobability sampling mentioned above, because of the interest in nonresponse bias rather than nonresponse rates, and because of the focus on semi-official surveys that should reflect the opinions and behaviour of the Dutch population, the remainder of this study will concentrate on nonresponse in random sample surveys.

2.5 *Survey mode and nonresponse*

When studying the effect of interview mode on response it should be taken into account that the response process will differ. In face-to-face surveys interviewers

can tailor their approach to characteristics of the dwelling and the first reactions of the person who opens the door. In a telephone survey there is also some interaction between interviewer and respondent, but the interviewer has much less information that can give guidance on how to proceed. In a face-to-face or telephone survey the decision to cooperate is generally taken during the introduction of the survey whereas the decision to complete a self-administered survey (mail, web) may be taken after inspecting the questionnaire ('stupid questions') or halfway through ('I am not going to finish this'). Breaking off interaction with an interviewer is generally more difficult and less common than not completing a self-administered questionnaire. In mail surveys it may be difficult to compute response rates and cooperation rates because one can often not be sure whether the respondents has received the mail or opened the questionnaire.

Partly because of these differences, the choice of a survey mode is generally a consequence of the topic of the survey, the length of the questionnaire, the type of questions, the characteristics of the target population and the available funds. In cross-cultural surveys such as the European Social Survey that will be discussed in chapter 10, for instance, it was considered important that all participating countries used the same interview mode to minimize mode effects, and face-to-face inter-viewing appeared to be the only interview mode that was feasible in countries with different rates of illiteracy and telephone and Internet penetration. Long questionnaires are generally not administered over the telephone, mail questionnaires are not sent to illiterate populations, and the elderly are not approached to participate in web surveys. This complicates comparison of response rates across survey modes. Furthermore, a distinction should be made between the recruitment mode and the interview mode. A face-to-face interview may be preceded by a telephone call in which the respondents is asked to cooperate. A self-completion questionnaire may have been handed out by an interviewer who obtained cooperation, gave instructions on how to fill in the questionnaire and came back later to collect the completed questionnaire. In some cases the mode of recruitment is left to the interviewer, or different modes are used for compliant and recalcitrant respondents. To compare response rates and processes across modes either meta-analyses should be used in which these different factors can be taken into account, or carefully controlled experiments (see De Leeuw, 1992).

In general it is assumed that face-to-face surveys receive higher response rates than telephone and mail surveys. Holbrook, Green and Krosnick (2003, p. 80) concluded from a comparison of (long) telephone and face-to-face surveys that the former are less expensive, have a shorter turnaround time and can be more standardized because of closer supervision of interviewers, but generally achieve lower response rates. There are exceptional cases with contradictory results, however. In a field experiment De Leeuw (1992) obtained lower response rates in a face-to-face interview (51%) than in both mail and telephone surveys (response rate 67%). Another exception to the rule is described by Reuband and Blasius (1996, p. 304) who achieved a 90% (!) response rate in a telephone

survey, and 71% in both face-to-face and mail mode. Interesting enough, the response rate for the face-to-face mode was 37% in a rather sloppily conducted first wave. Re-contacts about a year later with better-paid, more experienced interviewers yielded the additional 34%. The authors conclude (p. 315) that under favourable conditions response rates of 90% (telephone) or 70% (mail, face-to-face) should be possible.

Holbrook, Green and Krosnick (2003, p. 94, 113) ascribe the higher response in face-to-face surveys partly to the reassuring effect emanating from an interviewer on the doorstep: 'the nonthreatening and professional physical appearance of most interviewers and their equipment, along with their pleasant, friendly, professional, and nonthreatening nonverbal behaviors'. They also mention, however, that some respondents, especially women and the elderly, are reluctant to let a stranger into their home. In an attempt to obtain information from nonrespondents in the AVO1995 (see chapter 6) refusers were asked to answer a very short questionnaire (Van Leest & Burhenne, 1997). Only about a third of the nonrespondents complied, among whom the elderly and nonworkers were overrepresented and the more highly educated underrepresented. This also suggests that the elderly were less willing to let an interviewer into their home. Stocké and Langfeldt (2004) observed – to their surprise – that respondents whose last survey was face-to-face had more negative attitudes towards surveys than those who participated in other types of surveys. This effect disappeared, however, after controlling for the interviewee's direct evaluation of the survey with respect to the length and the necessary effort. This may mean that an often-mentioned advantage of face-to-face surveys (long questionnaires) may not be considered a benefit in the eyes of the respondents.

Combinations of modes within a single survey are possible, for instance when the face-to-face respondent answers a number on paper or the computer without the intervention of the interviewer. Very popular is the combination of a face-to-face interview and a drop-off questionnaire for self-administration. Increasingly popular is multi-channelling. This is strongly advocated by Voogt (2004) as a means of enhancing response rates, and can take two forms. Firstly, an inexpensive mode (a mail survey) among all sample units is followed by a more expensive mode (telephone survey) among those who have not yet responded, again followed by the most expensive mode (face-to-face survey) among those who could not or would not cooperate on the phone. This strategy was used by Diffendal (2001). Secondly, respondents can be offered the possibility to select their own favourite interview mode. An interesting variant was used in the Dutch Housing Demand Study (www.wbo.nl). Sample persons of whom a telephone number was known were asked a set of screener questions on the phone (the others were called on in person). If the screener questions indicated that the interview would be short, the interview went on over the phone. If the screener questions indicated that it might be a long interview, they were offered the possibility of having an interviewer come by for a face-to-face interview. In this case, more persons than expected preferred to complete the (long) telephone interview. They could also fill in

the questionnaire on the Internet, but that was not popular at all. The combinations of interview modes may be cost-effective and respondent-friendly, as is shown by experiences with the 'Neu Kontiv' design where the survey procedure has been made subordinate to the preferences of the respondents, with high response rates as one positive consequence (Snijkers and Luppes, 2000; Kalfs, 2001). It may well be the case that the possibility to influence the interview procedure is enough in itself to make the respondent favourably disposed towards the survey (see also section 5.3). Not imposing one interview mode may cause mode effects and measurement error, however.

2.6 *Differential sources of nonresponse: noncontact and noncooperation*

2.6.1 *The two parts of interviewer surveys*

Eligible sample persons may not participate in a survey because they have never been asked to cooperate or because, if asked, they decline to participate either because they do not want or are not able to participate. Accessibility (ease of contact) and amenability (likelihood of cooperation) are generally assumed to reflect two distinct dimensions on which sample households can be placed (Stinchcombe, 1981; Goyder, 1987; Groves and Couper, 1998; Lynn et al., 2002, p. 146). It is important to be able to classify nonrespondents into noncontacts and refusers, and respondents according to ease of contact and compliance, because:

- When trying to enhance response rates different measures apply to improving contactability and improving cooperation;
- When comparing surveys over time or across countries different nonresponse rates and a different composition of the nonrespondents (noncontacts and refusals) may be confounded with substantive differences;
- When estimating response bias or adjusting for nonresponse, knowledge about the underlying nonresponse mechanism (noncontact, refusal) should be available as contacting and obtaining cooperation are entirely different processes;
- When estimating response bias or adjusting for nonresponse, information on the difficulty of obtaining contact or cooperation is often used (see chapter 4) assuming that 'difficult' respondents are more like final refusers than easy respondents.

Making contact and obtaining cooperation in face-to-face surveys are two entirely different processes. Each contact attempt (or call [12]) has two outcomes only: contact or no contact. If first contact has been established, the contact phase is finished and the interviewer activity changes from making contact to obtaining cooperation. Before making contact, every contact attempt is basically similar (except for time and mode) and the prospective respondent will generally be unaware of unavailing calls [13]. With

12 Throughout this study and according to the literature a call is a personal visit to a sample address. In the study described in chapter 6 a call can also be a telephone call These will be identified as such, if possible (see section 6.2.2).

13 In practice this will be more complicated. Contact may be established with a household member and not with the designated respondent, the first contact may result in an appointment that may or may not result in an interview, and so on (see the next section).

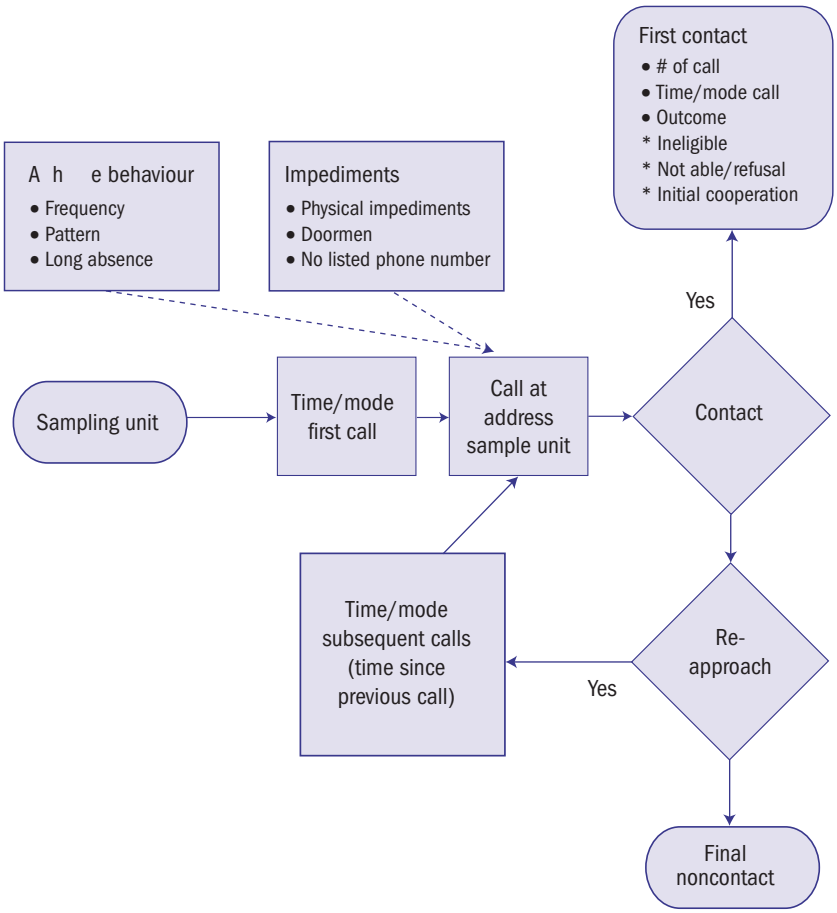
respect to obtaining cooperation, the situation is entirely different. The first contact can have three outcomes, namely an interview, the decision to treat the sample unit as nonresponse or the decision to try to make contact again and try to persuade the temporary refusal to participate after all. A second contact can also be seen as a prolongation of the interaction with the respondent in the first contact. This interaction builds up through subsequent contacts, and the decision to try to establish new contacts after a refusal is based on the combined outcomes of previous interactions. The different aspects of contactability and cooperation, and how to assess contactability and cooperation in practice, will be discussed in the next sections.

2.6.2 *Contact*

Establishing contact with sample units is the first step in obtaining response. If enough funds are available and the fieldwork period is long enough, a very high percentage of the population can be contacted in face-to-face surveys. In the European Social Survey 2002/2003, for instance, a noncontact rate of 3% seemed feasible in many participating countries (see chapter 10). Contactability is a major issue for two reasons. Firstly, if not enough time or money is available, it is important to know which bias noncontact might cause. And secondly, establishing contact may take an important part of the fieldwork costs. If contact can be established earlier, costs could be cut and, and there might be more time available for refusal conversion.

Contactability is not simply a respondent characteristic. Persons with a full-time job, for instance, may be found at home at the first call if the interviewer comes by in the evening, but may never be contacted if the interviewer sticks to working hours. Groves and Couper (1998, chapter 4) present a model on contactability in which it depends on physical impediments (locked apartment entrances), the at-home behaviour of the respondent and the interviewer call pattern. An adapted model, based on the fieldwork procedures in the AVO, is presented in figure 2.3.Ease of contact is often measured by the total number of calls made to an address. This is an imperfect measure of response propensity as it may both reflect difficulty of contact and reluctance, requiring further visits for refusal conversion purposes (see Lynn, 2002, p. 138). A better measure is the number of calls to first contact (Groves and Couper, 1998, p. 82) which comprises primary contact attempts only and excludes additional calls to contacted households that should be chalked up to reluctance to cooperate rather than being hard to reach. The number of calls to first contact is an imperfect measure too, as the timing of calls in face-to-face surveys is not randomly assigned but may be based on local knowledge of the interviewer or on information from previous calls (see the section below on call patterns). Nonetheless, it is generally acknowledged to be the best indicator of contactability. This measure is often not available as it requires detailed contact forms in which information on each individual call are recorded and keyed (see section 2.6.4). The outcome of the first contact may be that the sample unit is ineligible, a (temporary) refusal or unable, an appointment or a (complete) interview.

Figure 2.3 Model for contactability



Being at home

The at-home behaviour of the sample person is a function of household and individual characteristics which will be presented in section 3.2. These may also determine whether or not a listed phone number is available for the sample household, which may make contact easier. The likelihood of finding a respondent at home will depend on the amount of time a respondents spends at home, the at-home pattern and longer absences. The amount of time is an obvious predictor of contactability, the at-home pattern refers to a regular pattern of being at home or not at home, due to working times, school, weekly sports participation, etc. Blohm and Diehl (2001) point to longer stays in their country of birth as a reason why elderly Turkish people in Germany do not respond on surveys. Schnell (1997, p. 236) distinguishes persons who are hard to contact because they are away for a prolonged period (due

to a stay abroad or in the hospital) and persons who are hard to contact because they are not at home at the usual times (night nurses, persons who are very active in voluntary work). To reach the first, a prolonged fieldwork period is advisable, to reach the second a short fieldwork period may suffice in which calls are made at different times of the day. Persons who are away frequently or prolonged may still fairly easily be contacted if their family members are at home, who can tell when the person will be back or arrange an appointment.

Research generally focuses on the regular at-home pattern of respondents. Either general at-home patterns are studied (Groves, 1989) or independent information on the being at home of the sample unit is used in adjusting for noncontact, as in the Politz-Simmons approach. The latter approach would theoretically require a single call only to each sample unit as the results for persons not found at home could be estimated by incorporating the odds of persons being at home at specific times. This independent information is generally not available, however, and besides this is mainly an issue when respondent availability is the main problem (Smith, 1983; Ward, Russick and Rudelius, 1985; Groves, 1989, pp. 169-172; Smeets, 1995).

Call pattern

The second determinant of accessibility is the call pattern of the interviewer. To analyse the efficacy of call patterns and timing of calls, detailed call records are required. These are generally kept for telephone interviewing, but rarely for face-to-face interviewing. Earlier studies of call data (Campanelli, Sturgis and Purdon, 1997; Purdon, Campanelli and Sturgis, 1999; Groves and Couper, 1998, p. 101) give evidence that the contact rate on evenings is much higher than during working hours. Campanelli and her colleagues (p. 3-33) do not recommend, however, simply increasing the proportion of evening calls. This would make the success rate of each individual call higher and reduce the total number of calls, but would also most likely increase travel costs and not necessarily reduce the length of fieldwork, as the time available for making evening calls is limited. Interviewers may not be entirely susceptible to prescribed standard rules for call patterns, and are prone to vary these according to their own preferences and proven success strategies. It appears that interviewers generally appreciate being able to organize their own time schedule in making calls, to work during working hours and not to visit seemingly dangerous neighbourhoods during evening hours, despite the obvious success of evening calls. Groves (1989) also feels that it is not feasible to strictly prescribe call patterns.[14] The fact that individual prefer-

14 'Even if such estimates of conditional probabilities were available for all call numbers (e.g., probabilities of fourth call success at different times, given all possible combinations of times for the first three calls), it is unlikely that personal visit interviewers could take advantage of this knowledge. Their times of visitations are limited by concerns about the cost of travelling to a sampling area and thus the total number of sample houses that they can usefully visit at one time. Furthermore, the complexity of call procedures based on such empirical guidance would complicate their work.' (Groves, 1989, p. 99)

ences of interviewers have an effect on calling patterns and contact rates was shown by Lievesley (1983, p. 296) who found that busy interviewers had higher contact rates because: 'Interviewers with lower availability were calling at times when respondents were more likely to be at home, such as weekends and evenings.'

As mentioned above, the number of calls to first contact is an imperfect measure of contactability as interviewers may through local knowledge be aware of suitable moments to call on respondents and act accordingly. In a discussion the author had with interviewers who worked in the Westland region, the greenhouse market-gardening area of the Netherlands, it emerged that they had found out that the best time for contacting residents here was around noon, when the horticultural workers came home for lunch. And of course that is what they did.[15]

If interviewers develop their own calling strategies, based on their knowledge of the neighbourhood, characteristics that are ascribed to 'hard to reach' respondents may partly be due to particular call patterns. For instance, if interviewers are wary of making evening calls in inner-city neighbourhoods because they are concerned about their personal security, inner-city sample households will be hard to reach. This is not an individual characteristic, but occurs because they do not receive calls at times when the chances of success are high. Conversely, if local interviewers know that in certain neighbourhoods most people are employed or otherwise engaged during the day, they may start calling during evening hours and reserve their mornings for neighbourhoods where many elderly people live, and use the afternoons to visit residential neighbourhoods comprising families with children. This strategy is both very effective and likely to obfuscate differences in contactability.

Clues from the first, unsuccessful, call (comments from neighbours, children's bikes in the front garden, over-stuffed letterboxes) may also affect the timing of subsequent calls. Local knowledge, information from previous calls and interviewer circumstances may determine when sample units are contacted. These individual differences may be based on sensible choices and even according to instructions, but end up confounding the timing of calls and accessibility in nonresponse modelling (Groves and Couper, 1998, p. 101). This will certainly vex response modellers, as it fouls up their models of contactability because the interaction between household characteristics and interviewer call patterns may bias the estimates of contact probabilities for individual calls. Groves and Couper (1998, p. 82) lament that: 'In short, the ideal data set would have fully randomised visit times for all sample units – a practical impossibility', and end their treatise on the timing of calls wondering (p. 98) 'what characteristics of

15 They also got bags of sweet peppers and bunches of flowers from the horticultural respondents, which shows that handing over a bunch of flowers as incentive to respondents may not always be an obvious response enhancement strategy.

sample segments or neighbourhoods are related to interviewers choosing different times of day to call'.

Impediments

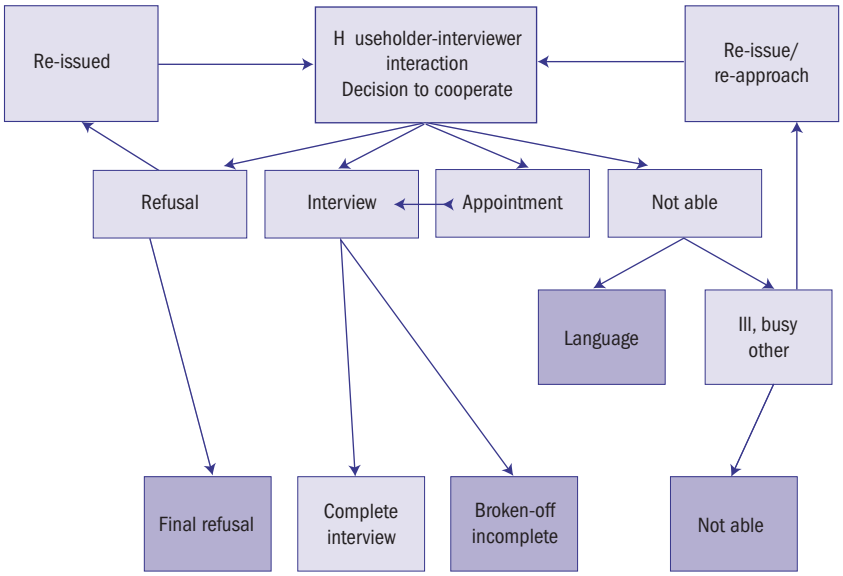
Apart from the timing of the call and the at-home behaviour of the respondents, physical impediments can obstruct contact between interviewer and sample person. Physical impediments play diverse roles in survey participation (see also Groves and Couper, 1998, p. 88). Firstly, they may simply make contact with the sample person impossible. Secondly, door grates, dead-bolt locks and alarm systems may indicate respondents who would not easily let an interviewer into their house and are probably reluctant to be interviewed. Thirdly, apartment buildings with locked communal entrances and intercom systems make face-to-face contact in the recruitment phase impossible, making it difficult for interviewers to show their ID-cards and a copy of the advance letter, and also to tailor their communication, and more easy for respondents to refuse cooperation. Besides, these impediments conceal whether a dwelling is empty or not, or that the resident is at home but not opening the door. And finally, the presence of a doorperson or gatekeeper may form an impediment. In surveys among senior housing residents in the Netherlands, access to the respondents was only possible with the permission of the staff. If contact may be made by telephone, the absence of a listed phone number is of course a very serious impediment as is the use of answering machines or voicemail devices as a screener for keeping out unwanted phone calls (see section 4.2).

Groves and Ziniel (2003) relate the presence of impediments (as observed by interviewers) to contactability in order to develop optimum call strategies. Using the presence of impediments in analysing contactability and reluctance requires that interviewers call at the house and register impediments in a standard way. If the recruitment is done by telephone, the information is not available. In the preparation of the European Social Survey it turned out that some participating countries refused to register physical impediments as this would be seen as snooping, and because this ran against their data protection and confidentiality laws. Other countries are less orthodox. In Canada, for instance, photographs were taken of sample household dwelling that were subsequently rated to determine the socio-economic status of the household (Goyder, Warriner and Miller, 2002, p. 3).

2.6.3 Cooperation

The model for cooperation in interview surveys in figure 2.4 is again derived from Groves and Couper (1989). Central is the interaction between interviewer and sample person, the moment when the survey is presented (topic, burden) and the prospective respondent decides whether or not to cooperate. This interaction occurs on the doorstep or on the telephone, and depends on survey characteristics and fieldwork design, the social environment, characteristics of the interviewer and individual

Fig re 2.4 Outcomes of sample person-interviewer interaction



and household characteristics of the prospective respondents. These factors will be discussed in the next chapters. The result of the interaction is a refusal, an interview, an appointment for an interview or the outcome that the respondent is temporarily or permanently unable to cooperate. After refusal the interview may be re-issued to the same or a new interviewer who will try to convert the refuser, the interview may be broken off and taken up again, additional material (drop-off questionnaires) may or may not be completed, appointments may result in an interview or not and respondents who were temporarily unable to participate (ill, too busy) may cooperate when they have recovered or are less busy. Figure 2.5 gives a more dynamic view of the fieldwork process after contacting a sample person, incorporating the outcomes of subsequent contacts and the decision of the survey organization whether or not to re-issue an interview. Both schematic models may be complicated by an intermediate phase of respondent selection within a household, and by the fact that an advance letter anticipates the doorstep information exchanges. Household, respondent and survey characteristics as determinants of the discussion to cooperate will be the subject of the next chapters. This section only outlines the process of trying to obtain survey cooperation and focuses on 'not able', refusal and refusal conversion.

Not able

Sample persons may not cooperate because they are simply not able to do this. A respondent who suffers from dementia or learning disabilities, who is illiterate or blind (mail survey), deaf (face-to-face or telephone survey) or does not speak the questionnaire language is not able to participate unless survey design characteristics are adapted. A respondent who is ill or otherwise engaged (birthday party, blocked drain) may be perfectly willing to cooperate later, if the fieldwork period is long enough. These causes of non-cooperation are not solely respondent characteristics, but are also related to design features. Admittedly, a selected sample person who was deceased by the time the interviewer called will not be able to cooperate (and still counts as nonresponse). In other cases, not being able may be an excuse for being unwilling to cooperate. In yet other cases, designating an elderly respondent as mentally incapable of being interviewed may be an easy way out for an interviewer afraid of a very long interview.

Not being able to participate may be a serious cause of nonresponse bias, for instance when ill people in a health survey cannot cooperate. Cohen and Duffy (2002, p. 21-22) conclude '... that prevalence of common sources of ill-health in the over 75's is likely to be underestimated, even by a carefully conducted health survey, but among the 'young elderly' such prevalence estimates are unlikely to be severely biased if a reasonably high response rate is achieved'. It might be added that the bias due to noncoverage will in many cases be much higher as – at least in the Netherlands – elderly persons in senior housing facilities are generally excluded from the sample frame in health surveys. In general, the percentage of sample units not being able to respond is generally very low, much lower than the refusal rate.

Refusal and refusal conversion

When contact with respondents has been established, and the respondent is obviously able to be interviewed, he or she may still refuse to cooperate. If so, reasons may be given for the refusal which may be used in the decision whether or not – and how – to re-approach an initially reluctant respondent, and in estimating bias. Reasons for refusal may be general or specific, and respondent-related or survey-related. Recorded reasons often refer to 'no time' (bad timing, too busy, no time), 'wrong subject' (not interested, don't know enough/anything about subject, too difficult for me, object to subject), 'surveys not useful' (waste of time, waste of money), 'surveys not pleasant' (interferes with my privacy /give no personal information, never do surveys, cooperated too often, do not trust surveys, previous bad experience). Of course, the reasons expressed for refusal may not reflect the sample persons 'real' opinion and just be a way of getting rid of the interviewer as soon as possible and avoiding discussions. The reasons given for refusal and their effect on bias will be discussed in section 4.1.2.

Figure 2.5 The process of obtaining cooperation after first contact has been established

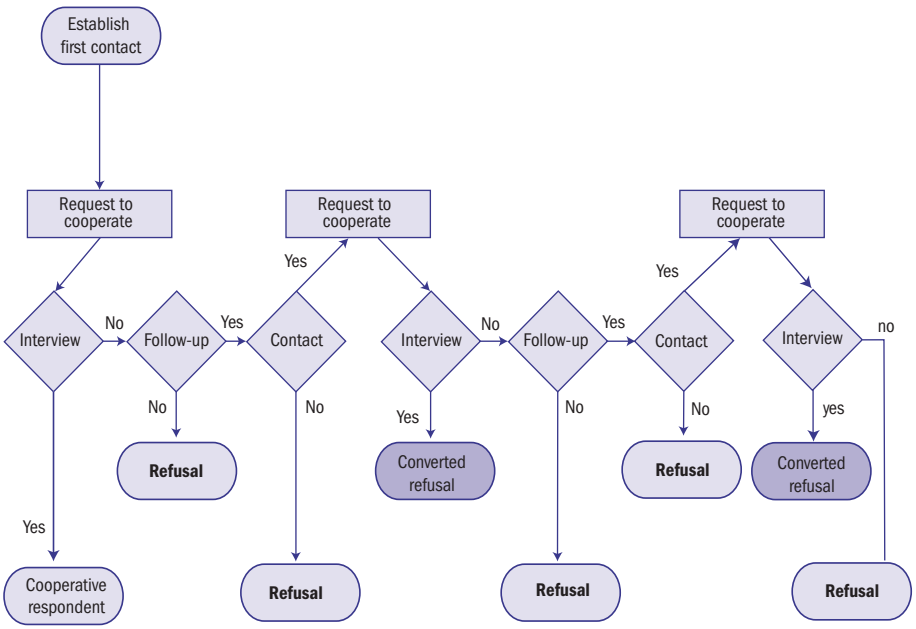


Figure 2.5 shows that a refusal may be permanent or temporary. Whether it is permanent or temporary will only be clear if the interview is re-issued and a new contact is established. The process of re-approaching an initial (hopefully temporary) refusal[16] and asking them again to participate in the survey is generally called refusal conversion. In some countries, refusal conversion is seen as bad practice. According to Däubler (2002, p. 24) German confidentiality laws forbid re-approaching a household after an explicit refusal. Refusal conversion can, however, substantially increase response rates. Schnell (1997, p. 190) concludes from his overview of the literature that of those who initially said they would not or could not cooperate 20 to 50 percent finally responded, and queries the existence of a hard core of refusers (see section 3.6). Lynn et al. (2002, p. 147) mention 10-20 percent for the United Kingdom. If refusal conversion is allowed, different strategies can apply. Either all refusals can be re-approached, which might result in uncomfortable situations in the field and serious complaints about the fieldwork organization.

16 Temporary refusals are also somewhat euphemistically called initially reluctant sample persons. One reason for this is that it is not always allowed to re-approach refusals (and many 'initially reluctant' sample persons cooperate in the end).

Or specific types of refusals may receive another visit, for instance those refusals who live on the route of a second interviewer, or those who refused to an interviewer who delivered poor quality in this survey. And finally, only those refusals may be revisited that have an acceptable chance of cooperating, according to the interviewer.

In countries where refusal conversion is considered good practice, the latter procedure seems to be favoured. Lievesley (1983) describes it as follows: 'How do we distinguish those people worth revisiting from those who are extremely unlikely to take part? Perhaps the most obvious way is to instruct the interviewers to record the reason for refusal and to make a subjective judgement regarding reissuing based upon this reason. ... reissuing on the basis of interviewers' assessments is preferable to random reissuing and we plan to continue experimenting with this procedure.' This strategy of re-approaching reluctant sample persons who seem likely to respond in the end is cost-effective and probably the only viable strategy. One hates to think what would happen if interviewers went back again and again to those sample persons who made it very clear that they absolutely did not want to cooperate, although subsequent chapters will show that it is possible to re-approach very persistent nonrespondents without unpleasant consequences, and that re-approaching promising cases only may increase response rates without improving the quality of a survey.

Deploying a new interviewer seems to be the accepted procedure in refusal conversion: '... refusal conversion efforts ... often entail the use of interviewers who are different in sex and age from the initial interviewer' (Groves, 1989, p. 218). New interviewers might be untainted by earlier refusals and, for whatever reason, appeal more to the sample person than the earlier one. Besides, new interviewers deployed in refusal conversion are often more experienced interviewers with a higher success rate. Finally, interviewers tend to see a situational refusal to them as a final refusal to cooperate in the survey. They may not be the best judges to decide whether or not someone who does not grant them in interview will cooperate in the future (see section 5.2). The original interviewers, however, might be more successful because they know more about the household to be contacted and the previous interaction. As Groves and Couper (1998, p. 266) say: 'Senior interviewers sometimes say that on refusal conversion efforts it is more important for them to learn what the householder said on prior visits than what the interviewer said.' So, the original interviewers might better be able to tailor their approach. Experienced interviewers may be rather successful at the second contact, because they do not push too much at the first contact and, by retreating at the right time, can tailor their approach and maintain contact about contacts so that they come back later, without exactly making an appointment the first time (see also Groves and McGonagle, p. 251). For this reason too, one might consider leaving refusal conversion to the original (experienced) interviewer.

2.6.4 *Distinguishing attainability and amenability*

Section 2.2 posed that rather than a general decline in response rates, there may have been a change in the composition of nonresponse, and that the share of noncontacts and refusals may be shifting. The previous discussion of contact and cooperation emphasized the general importance of distinguishing between ease of contact and reluctance to cooperate and the different processes that may cause noncontact and refusal. The next chapter will show that different groups in society may be more likely to be hard to reach or difficult to persuade to cooperate. What remains is to show how these groups can be distinguished. This requires paradata[17] and especially the registration of the outcome of each call (contact attempt) and each contact during fieldwork, reasons for refusal and assessment of future success. These are often jotted down by the interviewer in order to facilitate the data collection process, but not generally available in a systematic form to explore correlates of nonresponse (Groves and Couper, 1998, p. 49), at least not in face-to-face surveys.

Stinchcomb, Jones and Sheatsley (1981, pp. 374-375) end their analysis of nonresponse bias in attitude research by saying that '... every reputable survey should have as one of its variables describing every respondent a 'Temporary Refusal' variable, which is 1 if the respondent ever refused and 0 if the respondent agreed to be interviewed when he was contacted. In 1987 Goyder (p. 81) notes: While methodology researchers are by now conscious of differences between noncontacts and refusals, the accumulated literature remains disjointed because of neglect in explicitly portraying this distinction by analysing noncontact and refusal separately, but side by side with a common model.' Fifteen years later not much seems to have improved. Lynn and Clarke (2001, p. 9) point out the problem as follows: 'A limitation of our definition of hard-to-contact is that the data captured routinely on these surveys does not allow identification of interviewer visits made after contact was established at the address. We have only a measure of the total number of visits to each address (TNC). This is not ideal as a measure of difficulty of contact, particularly because successful interviewer strategies involve leaving and returning on another occasion, in order to avoid prompting a refusal (...). Thus, TNC is influenced by reluctance as

17 The major outcome of a survey are the *survey data*. This comprises the actual answers of respondents to surveys questions, including derived data such as scales, indicators and classifications and coded answers from open questions. *Metadata* describe a study, survey and fieldwork, ideally according to international data documentation standards. An excellent example of metadata of a specific study is the data protocol of the European Social Survey (ess.nsd.uib.no/2003/DataProtocol.pdf). *Paradata* give information on the fieldwork process, especially on the recruitment phase when contact is made and cooperation is sought, and possibly on the questionnaire answering process. Fieldwork information originates from administrative sources and from the interviewer (call records, call outcomes, reasons for refusal, observations of dwellings and neighbourhood, assessment of future success. Paradata can be recorded on paper or electronic forms, or on audio tape during the recruitment.

well as ease of contact.' One year later, Lynn et al. (2002, p. 136) utter a similar complaint: 'Thus, a major weakness of much previous research is that it either confounds ease of contact with reluctance or isolates one without considering simultaneously the effect of the other.'

Atrostic et al. (2001, p. 224) formulates quite detailed wishes: 'Likewise, it would be extremely helpful to capture case histories detailing outcomes each time a contact attempt was made differentiating whether it was to collect data or merely to make appointments. Retaining interim outcome codes would give researchers access to a wealth of information behind the broader noninterview categories currently retained, such as the number of contacts and whether each contact resulted in the same or a different type of outcome'. Duhart et al. (2001) struggle with seriously deficient paradata '... the lack of consistent and detailed interview case-history information results in ambiguities surrounding the conceptual and operational deficiencies of late/difficult interviews. For example, without knowledge of the number of contacts (attempted and made), how does one go about differentiating the truly difficult interviews from those that were mainly hard to reach?' and come up with critical recommendations: '... agencies should begin to routinely collect and retain case history information of interviewer contact attempts. This should include number of contacts and attempted contacts, interim outcomes that differentiate non-contacts from initial refusers, and final outcome codes that differentiate the same.'

Many survey researchers in countries other than the US and the UK feel that non-contact and refusal, and reluctance and ease of contact should be distinguished in nonresponse research, but that this information is generally unavailable. Schnell (1997, p. 150) observes the deficiencies in information on refusers and states (translated from the German): 'Firstly, we have the always incomplete contact forms due to negligence of the interviewer', and adds in a footnote (again translated from the German, p. 151) 'This is especially true for the Federal Republic, where hardly any importance is attached to the contact forms.' If the information is available, it is bound to contain errors, be inconsistent and incomplete and will only superficially cover what happened at each call. As mentioned above, Däubler (2002) mentions in his analysis of nonresponse in the SOEP panel that the confounding of the total number of contact attempts and number of calls to first contact may be not so bad in the German situation, as confidentiality regulations forbid re-approaching explicit refusers, which of course is no help in enhancing response rates. In the Netherlands, outcome codes are generally provided in fieldwork reports, but detailed paradata on each call are usually not collected, keyed or made available to survey researchers, with a possible exception of a Statistics Netherlands' survey analysed by Schouten (2003) and Bethlehem and Schouten (2003).

Kennickel (2000) gives an interesting overview of the pros and cons of different measures of 'lateness' or difficulty in the Survey of Consumer Finances and simul-

taneously highlights the importance of detailed paradata.[18] Bates and Creighton (2000) and Duhart et al. (2001) illustrate the operational difficulty of obtaining good paradata[19]: If a survey is considered a scientific endeavour, paradata have to be collected, keyed and analysed. If one is just interested in the substantive survey results, it might seem just a waste of time and money. However, if one wants to monitor and improve the survey process, and analyse, compare and enhance response rates, indicators of ease of contact and reluctance to cooperate have to be available, which require a meticulous registration of timing and outcomes of every call.

The general lack of detailed paradata appears to be a major impediment for nonresponse analyses and controlled studies of response enhancement. As collecting paradata is not the main purpose of a survey, sponsors, survey organizations and interviewers may easily consider these data of minor importance. Nevertheless, there are fieldwork organizations that regularly collect paradata and use detailed call records. In the preparation of the European Social Survey (ESS) an inventory was made of contact forms used by survey organizations in Europe and the UK, kindly forwarded by members of the International Workshop on Household Survey Nonresponse, making available material from the US, the UK, Denmark, Finland, Belgium and the Netherlands (Stoop et al., 2003). The development of contact forms for the ESS, their implementation and the information they yielded will be discussed in chapter 10.

18 'If all unresolved cases were worked uniformly over the field period, then elapsed time would be the best indicator of effort. However, many factors cause deviations from this condition. Interviewer attrition, under-staffing in some areas, avoidance of difficult cases, holidays, local variations, contractual constraints, and a variety of management decisions all contribute to a blurring of the relationship..... The number of attempts needed to complete a case may be a more direct measure of effort than the time measure. However, because all types of actions – negotiations with respondents, leaving a set of materials at a house where no one is home, mailing letters, leaving messages on an answering machine, etc. – have equal weight in this measure, it may overstate effort. If interviewers fail to act on cases where they believe they are less likely to succeed, attempts will tend to understate difficulty.' (Kennickel, 2000, p. 10)

19 'First, in 1999 the CPS instrument recorded contacts only for personal visit interviews (data on the number of telephone attempts and contacts is not currently captured). ..... The second, and perhaps more serious caveat to this indicator is the degree of between-interviewer variability with which it is recorded. For the counter to increase, the interviewer must access and open the case from the laptop, thus recording each personal visit made to the address. However, we have no way to confirm how often interviewers fail to actually open a case from the computer when visiting an address; that is, the interviewer could visit an address on several occasions, find no one home, but never activate the case prior to attempting contact. Such cases will understate the number of contacts.' (Bates and Creighton, 2000, p. 2

An equilibrium needs to be found between operational overhead and obtaining paradata. Or, as mentioned above, it should be made clear that paradata are a necessary part of a scientific survey and interviewers should be paid accordingly. This will not solve a problem observed earlier. In Germany re-approaching reluctant sample persons may run against privacy laws. In the ESS several Nordic countries said they could not provide information on nonrespondents (including reasons for refusal) nor make interviewers register information on the dwelling and the neighbourhood. In this case too, it should be clear that survey quality is not only determined by the quality of the questions or the response rate, but also by process control and detailed fieldwork documentation.

3 The usual suspects: hard to reach and reluctant to cooperate

3.1 Introduction

When discussing response rates in Europe it is generally acknowledged that survey participation of ethnic minorities is low. Several explanations can and have been suggested which serve as an example why identifying 'difficult' groups is more complicated than it might seem. Ethnic minorities more often live in big cities, which have lower response rates in general. They may more frequently live in apartment buildings and have unlisted phone numbers, both impediments for establishing contact. They might spend longer extended periods in their country of origin, which may make them harder to reach. There may be language problems that make interviewing difficult or impossible or that can only be solved by deploying native speakers who may not always be experienced interviewers. They may be less used to filling in forms, more often have a lower education, lower incomes and poorer jobs, and thus might share their low response rates with other persons with a lower socio-economic status. They may not identify themselves with the sponsor of the survey, or be wary of government involvement. They may not be willing to talk to a female interviewer, or to a male one. All of these factors may be true to a certain extent. If so, the cause of the low response rates is not that they belong to an ethnic minority. That would just be a correlate of nonresponse. But they still would be a problematic group in survey research. To find out what are the causes of low response rates, to enhance response rates and to estimate nonresponse bias among ethnic minorities, one should keep apart socio-structural characteristics and attitudes, accessibility and amenability, and respondent characteristics and survey characteristics.

According to Goyder, Lock and McNair (1992, p. 39) 'Studies of nonresponse on surveys too often are conducted within watertight compartments. One group of authors investigates the socio-demographic profile of nonrespondents. Another conducts experiments with field procedures on topics such as cash incentives, characteristics of interviewers, or topics of survey. A third examines geographic aggregate variables such as city population size or density. These studies are difficult to synthesize.' It is especially because of the drawbacks of these isolated approaches that the conceptual model of Groves and Couper (1998) is so useful, as it keeps different factors apart, and acknowledges for instance that the socio-demographic composition of the population differs between urban and rural areas, incentives may work better for rich or for poor people, the topic of the survey may be relevant for some and irrelevant for others, population density may have a different effect on contact rate and population rate, and some persons may simply not like surveys and answering questions.

Section 2.6 presented the two different components of the difficulty of achieving an interview: ease of contact (possibly resulting in noncontacts) and reluctance to cooperate (possibly resulting in refusal). Although this section focused on these components *per se*, it was unavoidable to sometimes indicate which groups might be particularly difficult to contact or to persuade to partake in the interview. The present chapter identifies these difficult groups while being aware that many household and individual characteristics are related and that identifying difficult groups ignores the intrinsic causes of response behaviour that may be related to socio-structural indicators and survey characteristics. The latter factors will be discussed in chapter 4. In the present chapter characteristics of respondents and nonrespondents and their neighbourhoods will be presented as identified by the literature, not as causes of nonresponse but because they are frequently mentioned, are easy to identify and easy to observe by interviewers, and may determine interviewer behaviour.[1]

More than 25 years ago, the Dutch market research group Inter/view-groep (1978) published a study on nonresponse aimed at generating a discussion on response enhancement procedures and at greater transparency in nonresponse reporting. At that time, a response rate of 95% was considered to be feasible. The percentage of persistent refusers was smaller than 5%, and refusal conversion attempts by another interviewer were successful in 80%. Accessibility was related to family composition (low if the 'housewife' had a paid job, for one and two person families and for sample persons aged 35-49 years, high for mothers of young children and persons aged 65 years or older). Amenability was negatively related to age and urbanicity, and positively to social class. The contact rate was somewhat higher during evening hours; respondents preferred afternoon interviews. The attitude towards surveys was positive, although converted refusers less fancied participating themselves. Converted refusers in this survey were somewhat less affluent (possibly related to age) and more often had a dog.

This early study distinguishes between those groups in society that are hard to contact and those that are reluctant to cooperate. The present chapter makes the same distinction. Section 3.2 will focus on those groups in society that are hard to reach because of practical impediments and the reduced likelihood that a member of the household is at home. Section 3.3 gives an overview of factors that are often mentioned in relation to survey cooperation, namely demographic factors, education and form literacy, socio-economic status, urbanicity and ethnic minorities. As mentioned above, these factors cannot be seen in isolation.

1 There is anecdotal evidence of interviewers not pushing the doorbell of sample addresses with aspidistras on the windowsill because this indicated elderly people and possibly very long interviews.

3.2 Hard to contact

Throughout the years the same factors have been identified as determining ease of contact. Smith (1983) mentioned the following correlates of being found at home: labour force participation, socio-economic status, life stage (age and marital status), health and sex. According to Goyder (1987, p. 84) socio-economic status, as indicated by education, income or occupation, is generally found to be negatively related to accessibility. Singles, workers, apartment dwellers, big city dwellers and high occupational SES groups are more difficult to contact, and the elderly, larger families and the residentially stable easier

From their analysis of interviewer call strategies and neighbourhood characteristics Campanelli, Sturgis and Purdon (1997, 3-13) concluded that 'it is more difficult to make contact with address residents in areas of high population density, in areas where a high proportion of the dwellings consist of flats, and in areas where there is a high proportion of persons in non-manual occupations. In contrast in areas with a higher than average number of two person households making contact is easier.' Groves and Couper (1998, p. 115) found that households with young children or elderly adults (both outside the labour force) were more often at home and more easily contacted. In addition, households with more adults are easier to contact than those with fewer adults, probably because of the increased probability of finding one household member at home. They assumed that the lower contactability in urban areas might be related to the different household composition and possibly to the restricted entrance to apartment buildings. Lynn et al. (2002, p. 142) analysed ease of contact for a number of British surveys. Difficult to contact households were more often in employment, younger and less often white than easy to contact households (both easy to contact and cooperative). Lynn and Clarke (2001) concluded from an analysis of British surveys, that hard-to-contact persons are more likely to be regular smokers and heavy drinkers and less likely to have a long-standing illness, and more likely to have lower blood pressures and body mass index. This might be due to hard-to-contact persons being both younger and more likely to be employed.

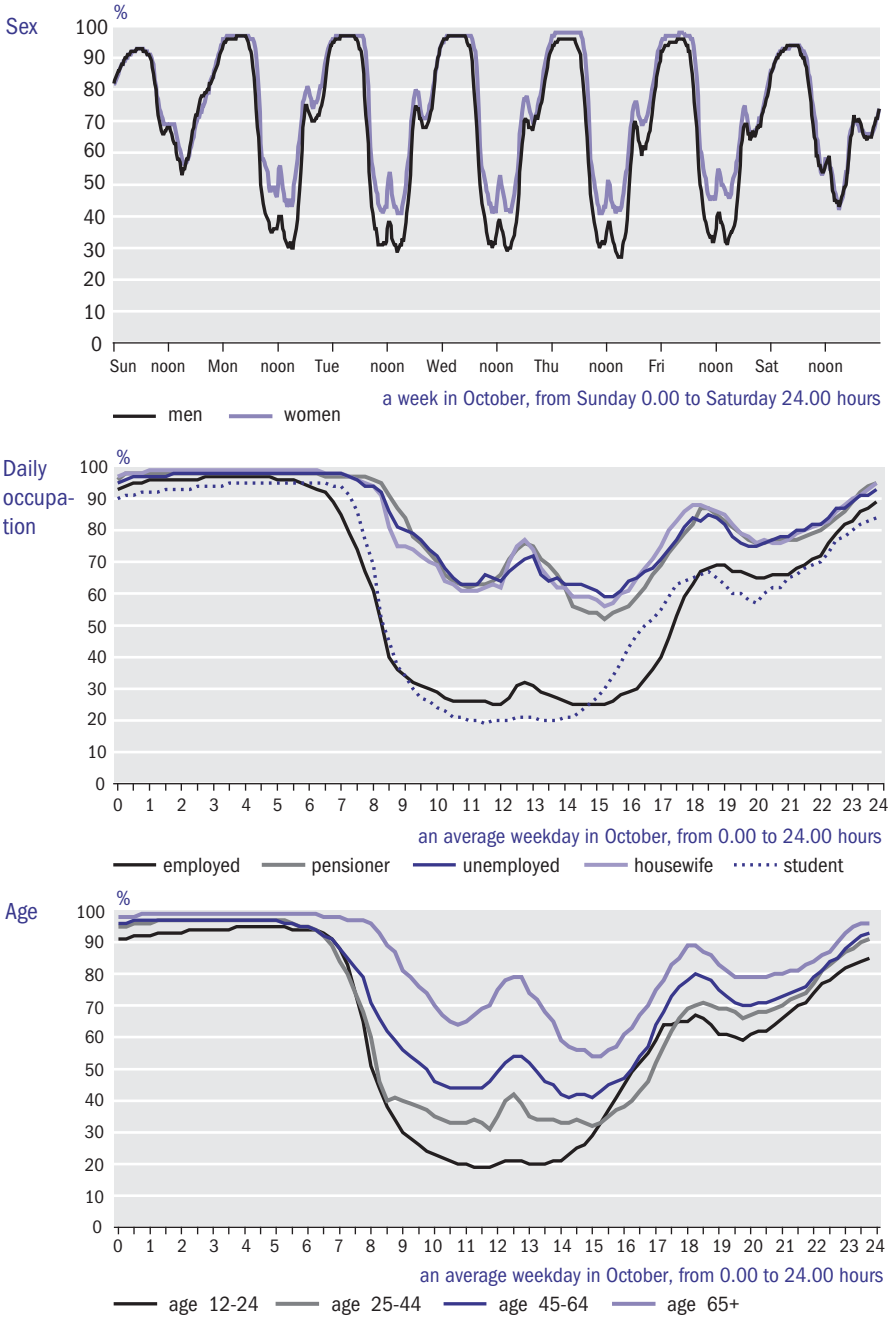
The studies above indicate that accessibility of sample households is a direct function of the probability that a household member is at home, which is related to the number of household members and their daily occupation. The best measure of being at home would be a direct estimate from respondents. Japec and Lundqvist (1999) analyse the at-home pattern from respondents in the Swedish Labour Force Survey 1996 who were asked a number of questions regarding their at-home-pattern for a certain day. It turned out that big city respondents had a different at-home pattern than other respondents and that weekends seemed especially good for finding people at home. A second best measure on being at home is derived from time use data from independent surveys, as presented in figure 3.1 for the Netherlands in a week in October 2000. Figure 3.1a shows that that women are often more at home than men and

that on weekdays around noon about 30% of the women and 40% of the men will be found at home, compared with 70% after 6.00 p.m. At weekends, and especially on Sundays, the likelihood of catching a respondent at home is larger. Of the old age pensioners at least 50% are at home at any given time, whereas less than 20% of the youngest age group will be around at noon on weekdays. In figure 3.1b old age pensioners, housewives and the unemployed have a very similar at-home pattern and can hardly be distinguished. They differ a lot, though, from employed persons and students; during weekday mornings and afternoons only a quarter or a third will be found at home. Being at home increases with age (figure 3.1c). The patterns in figure 3.1 indicate how accessibility can be improved: make calls in the evening or during weekends. The analyses of Campanelli, Sturgis and Purdon (1997, p. 3-33) confirm the efficacy of these call patterns (see section 2.6.2).

Section 2.6.2 also mentioned that persons may be hard to contact because they are frequently not at home when the interviewer calls or because they are away for prolonged periods. This may concern elderly persons from ethnic minorities. Blohm and Diehl (2001) point to longer stays outside the holiday periods in their country of birth as a reason why elderly Turkish people in Germany do not respond to surveys. This might also be relevant for old age pensioners in the Netherlands who hibernate in sunny southern European countries.

Physical impediments may interfere with contacting sample persons. Not many private addresses in the Netherlands are part of gated communities, nor do apartment buildings have doormen who regulate entrance. An intercom system is not necessarily an impediment to establishing contact but may render obtaining cooperation more difficult. Telephone surveys or face-to-face surveys in which respondents are recruited by telephone have to contend with other barriers to securing contact with prospective respondents. Screening devices and answering machines make contacting respondents difficult. In the Netherlands, people who lack economic, and/or cultural, and or social 'capital' are overrepresented among those without a telephone or with an unlisted number (Van Goor and Rispens, 2004). Telephone interviewing becomes more difficult as fixed telephones are replaced by mobile ones (De Leeuw et al., 2002). Mobile phones may also have advantages. Kuusela and Simpanen (2002) list a wide range of potential effects that might be beneficial or harmful for telephone surveys. One advantage is that notoriously difficult groups such as young single men might be easier to reach by mobile phone, as are people on a holiday (and in the mood for an interview). Another advantage is that by now short SMS surveys are possible. Holbrook, Green and Krosnick (2003, p. 94, 110) mention the well-established fact that low-educated, low-income and minority respondents are underrepresented in telephone samples. The studies they mention suggest '... that telephone respondents are more likely to be well educated, to have high incomes, and to be male, older and white .... This is further reason to expect that socially vulnerable groups will be less well represented in telephone surveys.'

Figure 3.1 Percentage of population at home (a) at a given time, according to sex; (b) on an average weekday, according to daily occupation; (c) according to age group



Source: SCP Time Use Survey 2000

3.3 Reluctant to cooperate

3.3.1 Out of the researcher's control

About ten years ago, Dutch senior survey researchers identified the following groups as nonresponse-prone (Kalfs & Kool, 1994): younger persons, older persons, persons with a lower education level, urban dwellers and ethnic minorities. Groups like these are often identified as difficult groups without always properly distinguishing between contact and cooperation, and between determinants and correlates of nonresponse. In this section those socio-structural or background variables will be identified that are frequently mentioned in relation to survey cooperation. Background variables are generally not the topic of a survey, although the distinction between background and core variables is not always clear: in a health survey, type of employment is a background variable, whereas in a labour market survey it is a core variable. Whether or not a particular group is generally less likely to cooperate in a survey is not under the researcher's control (Groves and Couper, 1998, p. 30). What a researcher should do is try to find out why and how these background variables are related to survey cooperation. This may be relevant for possible nonresponse bias, but also for adapting fieldwork procedures or survey characteristics that are under the researcher's control and that may make survey cooperation more attractive to these difficult groups. These field strategies will be discussed in the next chapter.

In this overview of the literature, it will not always be possible to keep contact and cooperation apart. In some cases, contactability will be mentioned intentionally, because of the conflicting influence of background variables on contact and cooperation.

3.3.2 Age, sex and family composition

Younger persons are less often at home and therefore more difficult to contact. Elderly persons may be more socially isolated and thus have lower cooperation rates (Groves and Couper, 1998, p. 133). A higher incidence of health problems may make the (very) elderly physically or mentally unable to participate in a survey (Cohen and Duffy, 2002, p. 21-22). Groves and Couper (1998, p. 133) expect and find a mixed effect of age from an inventory of the literature on age and overall response rates. With respect to refusal, they find support for less cooperation from the elderly. This effect disappears, however, when controlling for household size. The failure to find the expected effect might be due to conflicting influences, for instance an increased fear of victimization among the elderly might be counteracted by a higher sense of civic duty towards government surveys. Including neighbourhood characteristics, interviewer characteristics and topic salience did not bring about the expected age effect, however. In their multivariate analyses, they even find a curvilinear effect of age on cooperation rate, where both young and old households cooperate more (p. 148, 150). The authors ascribe this to a higher interest in social participation among the younger households, more curiosity about efforts to seek information from

them, and more experience with standardized information-seeking associated with schools and jobs, and stronger norms of civic duty among the elderly. Anecdotal evidence from the Netherlands seems to confirm this: according to interviewers students are very difficult to reach but interested in answering questionnaires, as that is related to their study. In section 1.3 the beneficial effect of the norm of civic duty in the Netherlands was questioned, but the presence of this norm among elderly people is illustrated by elderly persons giving notice of the decease of prospective respondents through sending copies of death certificates in a non-compulsory survey.

Men often have lower response rates than women. This will partly be due to the fact that they are more often not at home (figure 3.1a). On the other hand, men more often have jobs, and thus may be more used to filling in standardized forms, whereas women might be less eager to have a stranger in their home. The latter fact might be especially important for elderly women. A complicating factor in the importance of sex is that survey cooperation might be a family decision, even when one particular person is the designated sample unit. One reason for this is that the interviewer may have to negotiate on the doorstep, before it is known who the sample person is. Refusal by proxy may not always be distinguishable from a direct refusal. Furthermore, even if the right person has been selected and contacted, it may well be that other household members object to the interview.

The presence of children seems to increase survey cooperation. Larger households are easier to contact than small ones, and it may also be easier to obtain the cooperation of a large household in a survey where every responsible adult can function as a household informant. In this case larger households present a larger substitution pool, increasing the likelihood that at least one person will cooperate. Groves and Couper (1998, p. 123) mention only one exception, namely the British Expenditure Survey that stood out because all adult members of the household had to participate. In multivariate analyses only the difference between single person households and other household types remained. Koch (1993) found in an analysis of response rates on the German ALLBUS over time, that the expected increase of single-person households in Germany was not reflected in an increase of single-person households among survey participants, most likely due to increasing non-contact and refusal among singles.

The striking position of single-person households warrants a closer look into the composition of this group in the Netherlands in 2000. Figure 3.2 shows that male (45%) and female (55%) singles are clearly very different. The number of male singles has a peak around 29 years of age and then gradually diminishes. The number of female singles has two peaks, at around 25 years (somewhat less pronounced than the male peak) and a very large 'widow's peak' of elderly women in their late seventies. Of the single women, 38% are aged below 50 and 62% 50 or older (71 years on average), whereas the composition of single men is exactly the opposite as 69% are below 50

and 31% 50 or older (65 years on average). Because of these differences in sex, life course and age-correlated issues such as education and employment, 'singles' are not a single, homogeneous group. The image of singles as swinging yuppies is thus entirely wrong, if only because only 19% of the female singles and 23% of the male singles live in big cities.

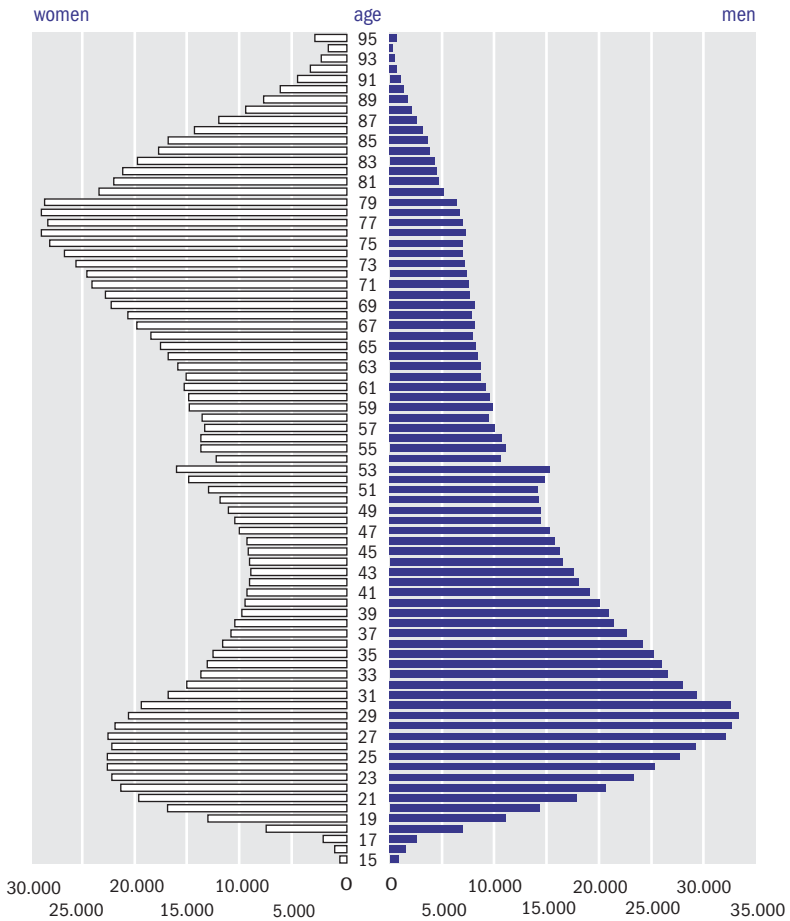
3.3.3 Education, form literacy and socio-economic status

Several studies have highlighted the relationship between educational level, form literacy and willingness to participate in a survey. Brehm (1993, 31) suggests that interviews would be seen as tests and thus more attractive by persons with a higher level of education. He found, however, an overrepresentation of persons with a lower education. Groves and Couper (1998, p. 128) hypothesize that the better educated have profited from earlier form-filling efforts and thus may be more prone to cooperate. They point to the consistent finding from the literature that the lower educated groups more often fail to participate in surveys, but find no such results in the governmental surveys they study. They also discuss the possible benefits of cooperating in a survey (p. 122) such as the enjoyment of thinking about new topics, and the possible costs, such as the cognitive burden incurred in answering and comprehending the survey questions. As the cognitive burden of cooperating in a survey might be less for the more highly educated, this might result in a higher cooperation among the higher educated (see also Tourangeau and Smith, 1996; Holbrook, Green and Krosnick, 2003, p. 82).

Socio-economic status (SES) is a household characteristic that comprises several aspects, namely education (related to form literacy), income and housing costs (related to urbanicity), occupation and employment (related to discretionary time). Being employed (or following a study) are clearly reasons for being difficult to contact and thus for nonresponse. Groves and Couper (1998, p. 126) discuss the possible relationships with household SES in the context of the social exchange theory. Either lower SES groups cooperate better because they use public assistance, whereas higher SES groups just resent government intrusion. This would cause a negative relationship between SES and cooperation. Or, lower SES groups feel they are unjustly disadvantaged and high SES groups might feel they are always targeted for contributions and have to pay high taxes to take care of others. This would suggest a curvilinear relationship. The results they cite give no clear evidence, one way or the other. In their own analyses they find evidence (p. 152) that lower SES groups (as measured by housing costs) show a higher cooperation propensity. The tendency for those in high-cost housing to refuse survey participation may be due to the fact that (in the US) residences in urban, high-density areas are relatively expensive.

Goyder (1987, p. 83-85) concluded from both the literature and his own analyses that in the US occupational SES and response propensity are highly positively correlated. Upper strata are less accessible, but show a higher response following contact. Goyder,

Figure 3.2 Sex and age single persons households, 1 January 2000



Source: CBS Statline 2004-02-08

Lock and McNair (1992) distinguished between individual SES and geographic aggregates. Home ownership and property value were positively correlated with response. Goyder, Warriner and Miller (2002) estimated SES by taking photographs of the dwellings of sampled households and having these valued by real estate agents. Their mail survey had a substantial SES bias, as high strata were easier to contact (less mail undeliverable), responded earlier and responded better to follow-up mailings and reminders. Te Riele (2002) found an underrepresentation of the lower strata in household surveys from Statistics Netherlands. Indicators of overall nonresponse are lower education, ethnic minorities, lower income, renters and inhabitants of apartment buildings. These groups are also predominant in big cities.

3.3.4 *Urbanicity and ethnic minorities*

The majority of studies agree that big city dwellers are harder to contact and more reluctant to cooperate. According to Couper and Groves (1996) and Groves and Couper (1998, pp. 176-187) urbanicity is the main social environmental factor on survey cooperation, and cooperation decreases as urbanicity increases. They ascribe the impact of urbanicity to higher population density, crime rates and social disorganization, three indicators of social cohesion. In their analysis it turns out that these variables are highly correlated, and that their combined effect is reduced by adding household-level controls, but a large part of the effect remains. Jansma, van Goor and Veenstra (2003) analysed neighbourhood differences and survey cooperation in a telephone survey in the Netherlands. They found a negligible effect of neighbourhood characteristics after incorporating socio-demographic household level variables. The effect of the latter was also very small, however.

The lower social cohesion in urban areas might result in less trust and greater fear of letting strangers into one's home. Lack of trust in strangers, privacy concerns and fear of government intrusion might also be indicative of social isolation and thus affect responsiveness to a survey (Brehm, 1993, pp. 52-56). Dillman (2000, pp. 19-21) highlighted the importance of trust and outlined ways to establish trust in order to enhance response rates. Holbrook, Green and Krosnick (2003, p. 94, 113) ascribe the higher response in face-to-face surveys partly to the reassuring effect emanating from an interviewer on the doorstep (see section 2.5) but also observe that some respondents, especially women and the elderly, are reluctant to let a stranger into their home. Fraboni et al. (2002) analysed the multiple effects of urbanicity, respondent characteristics and interviewer characteristics in an Italian survey, and found a high risk of refusal in the case of the elderly, people living alone, people living in large towns, people in the northern part of Italy, and people contacted by students with little interview experience. Urbanicity variables had the largest effect. They interpreted the combination as confirming their hypothesis that fear of opening the door to strangers decreases cooperation.

Urbanicity may also be negatively related to survey cooperation because of the different composition of the urban population. The Dutch big cities (Amsterdam, Rotterdam and the Hague) are generally seen as problematic areas in survey research. As early as 1977, Van Tulder pointed out the low response rates in Amsterdam, and the negative attitude towards survey research of a large part of Amsterdam's inhabitants (Rotterdam was quite positive). Table 3.1 illustrates the relationship between urbanicity and ethnic minority groups and income in the Netherlands. Dutch neighbourhoods with high percentages of ethnic minorities are characterized by more rented homes, cheaper owner-occupied properties, fewer detached and terraced houses, lower incomes and more unemployed (see for instance Van der Wouden en de Bruijne, 2001).

Table 3.1. Population, ethnic minorities and income in the Netherlands (2002)

	total population	ethnic minorities (N)	ethnic minorities (%)	non-western population (N)	non-western ethnic minorities (%)	deviation standardized incomes from national average (2000)
Amsterdam	731288	324560	44	227330	31	-3
The Hague	441094	176910	40	122940	28	-10
Rotterdam	592673	237040	40	180650	30	-3
The Netherlands	15863970	2775325	17	1408790	9	

Source: Statline (SBG2002/3) en CBS-webmagazine 3rd February 2003 (www.cbs.nl)

Immigrant groups in the Netherlands are nowadays generally perceived as less likely to respond than the indigenous population. Voogt (2004, p. 106) states that they are generally concentrated in specific areas, and are less integrated and possibly less susceptible to invitations to participate in national surveys. However, their circumstances were similar – or worse – 20 years ago whereas a comparison of 20 years of surveying among ethnic minorities shows a radical change in response rates over time. The 1984 quality of life survey among Turkish and Moroccan inhabitants (CBS, 1987) resulted in a nonresponse rate of 24% among Turkish and 36% among Moroccan inhabitants. The first outcome, especially, was low compared to other CBS surveys. In addition, the health survey of Turkish inhabitants 1989/1990 (CBS, 1991, p. 22) had a response rate of 83%, which was substantially higher than in similar surveys among the native Dutch population (Health Survey 1990: 58%). Bronner (1988, pp. 42-43) observed at that time: 'The Dutch score higher in all categories for non-cooperation: they are ill more often, they are out more often and have a higher rate of refusal' (compared to Moroccans and Turks). In his study a more integrated group such as the Italians were very similar to the Dutch. De Leeuw and Hox (1998) explain this by the scarcity principle: the more often a group is asked to participate in a survey, the less special this is and the lower their participation. They expect the cooperation of minority groups to decline over time, which is what happened with the response rates of the survey on the Social Position and Amenities Use of Ethnic Minorities (SPVA), co-sponsored by the SCP (www.scp.nl/miss/spva.htm). From this point of view nonresponse, at least in the Netherlands, can be considered as an indicator of integration.

As immigrant groups represent a large share of the inhabitants of European countries and form a group with specific characteristics and specific problems, the high nonresponse rate among this group is a reason for concern. Schoen et al. (2000, p. 16) concluded that ethnic minorities were substantially underrepresented in a telephone victimization survey, and especially those with a poor education and out of work. Schmeets and Janssen (2002, p. 7) found an underrepresentation of ethnic minorities in a face-to-face election survey run by Statistics Netherlands, but here – contrary to

expectations – ethnic minorities with a low socio-economic position, measured by dwelling value, were not less likely to cooperate. They did not distinguish between noncontact and refusal. In a similar study on election data Schmeets and Michiels (2003) again found an underrepresentation of ethnic minorities, this time especially those in urban environments. These results show that in the Netherlands the higher nonresponse among ethnic minorities cannot be ascribed solely and uniformly to their socio-economic position. This is contrary to the study of Groves and Couper (1998, pp. 148, 152) on the effects of race and ethnicity on cooperation. They conclude that the apparent tendency of racial and ethnic minorities to cooperate less with surveys is largely explained by their socio-economic status.

Blohm and Diehl (2001) report on a study of the survey participation of Turkish migrants in Germany. They conclude that response rates can be enhanced by using translated questionnaires. Native interviewers make interviewing in minority l anguages possible and thus rule out nonresponse due to language problems. These native interviewers may not always be the most experienced interviewers, however, which may have a negative impact on response rates and survey quality in general. Contrary to expectations, they did not find Turkish women especially hard to interview. They were easier to contact, were more ready to cooperate and male interviewers obtained even better results than female interviewers. The main reason for the lower response rates among the elderly Turkish migrants was noncontact, possibly because they spent longer periods in their country of origin, even outside the summer holidays.

3.3.5 Opportunity costs

Time concerns may be an important impediment to survey cooperation, either because the interviewer asks for cooperation at an awkward moment or because the sample person says or feels that being interviewed takes too much time. This might be particularly true in interviewer surveys and less so in self-completion surveys where the respondents can answer the questionnaire in their own time. Of course, in interview surveys the interviewer can also call again at a more suitable moment if the first call was not convenient. A lack of discretionary time might be a more permanent impediment to survey cooperation: 'All other things being equal, the burden of providing the interview is larger for those who have little discretionary time. Time limitations of the household should affect both contact and cooperation' (Groves and Couper, 1989, p. 122). The authors did not however find much evidence that reduced time at home leads to reluctance to cooperate with surveys (p 124). This might be due, as they say, to the fact that they do not have complete indicators of time spent away from home.

If time constraints are an important impediment to obtaining cooperation, long questionnaires ought to receive less cooperation than short questionnaire. The first question that comes up is, of course, what is long for a questionnaire. Dijkstra and

Smit (2002) presented evidence that a 10-minute telephone interview obtains higher response rates than a 15-minute interview. Long ago, Becker (1972) conducted eight-hour interviews in an institutional setting. He observed that, 'generally speaking, interviews of more than three hours are rarely conducted if the respondents have to give their cooperation in their off-duty time' (p. 160). Dillman (2000, p. 305) suggests that a questionnaire may even be so short that it stops making sense to respondents. Bogen (1996) has studied the relationship between interview length and response rates and finds the evidence inconclusive. Stocké and Langfeldt (2004, p. 9) assume that face-to-face interviews may be less liked because they can take longer than other survey modes. Loosveldt and Storms (2003) also observed a more negative attitude towards 'this particular survey' as the perceived (rather than the actual) survey length increased. It has also happened that respondents appreciate the interview so much that they invite the interviewer to come back, or become so involved in the topic that the interview turns into a lengthy discussion on topics in which they initially showed only disinterest and reluctance.

Within the opportunity cost framework one would expect that the same busy individuals who are hard to reach would be loath to spend their time on answering a questionnaire. This would make the amount of discretionary time a predictor of survey cooperation. The amount of discretionary time is an individual or household characteristic related to several socio-structural characteristics. In larger households, housework may require more time but can, in principle, be shared by more persons. Small children require childcare. Employed persons spend time on work and commuting, school pupils and students spend time at school or university and studying, etc. Even voluntary work may bring obligations that leave little time for survey participation. From this point of view, nonresponse in time use surveys is especially interesting. In a recent issue of *Loisir & Société*, devoted to time pressure, stress, leisure participation and well-being, a round table section was devoted to the relationship between participation in time budget studies and being busy (Zuzanek and Veal, 1999, pp. 547-582). Zuzanek (1999, p. 547) introduced the section with the arguments that can be raised in discussing the value of time use surveys: 'busy people do not have the time to complete time diaries and, as a result, are under-represented in national time use surveys. According to the counter-argument, busy people find time for all sorts of things, and are more likely to respond to time-diary questionnaires' Goyder (1987, p. 86) gives evidence for the latter hypothesis. He states that, although gainfully employed workers are elusive for fieldworkers to contact '... Non-workers perhaps command more free time for answering surveys but, unless the topic is highly salient to their own circumstances, they may also feel little involvement in many topics examined in survey research. Response seems particularly low among the unemployed, as distinguished from those voluntarily absent from the labour force'.

The contributions to the 1999 *Loisir & Société* issue generally contradict the assumption that busy people cooperate less in time use surveys.[2] Robinson introduced the Newtonian model of behaviour into time use research 'Bodies in motion stay in motion, while those at rest stay at rest' and cited Parkinson's (1958) famous law: 'If you want something done, give it to a busy person'. He pointed out that more active attenders in live arts activities are more active in other away-from-home activities (like sports or movies) than non-attenders, and worked longer hours, and also that the amount of time spent on an activity is not necessarily related to the output. He saw no indication that respondents were less active than nonrespondents, and found watching TV to be more popular among survey dropouts (even if corrected for age, education and working time) which indicates that it is unlikely that time pressure was the cause of dropping out.

Knulst and Van den Broek (1999) compared the final sample of Dutch time use surveys with the general population. Surprisingly, the least busy groups were underrepresented, i.e. people not involved in paid work and not burdened with high combined loads of paid work, education, childcare and other household obligations. The authors concluded that there is no evidence that busy persons cooperate less in time use surveys. An alternative interpretation is that young and older respondents are underrepresented in the survey; these are also the people who have fewer obligations in the sphere of paid work, childcare and education (for the elderly). The authors show that males, middle-aged persons, main wage-earners and the (self-)employed are busier than others. Rather than concluding that the least busy groups are less likely to participate in a time budget survey, it appears that the young and the old may be both less busy and less likely to cooperate. They may have had their own reasons for not agreeing to keep a detailed time budget diary for an entire week. Within age groups there is no clear-cut relationship between survey participation and busyness.

Pääkkönen (1999) was able compare respondents and nonrespondents in a time use survey with data from an initial interview, which contained questions on working hours, leisure participation and subjective sense of time pressure. The dropouts did not feel more rushed, nor experienced more stress. Working hours had a negative effect on survey participation only among the self-employed, and working overtime had no effect. Feeling rushed at work had a negative effect on survey participation, but the working dropouts '... considered the hurry to be less stressing than those who filled the diary out' (p. 578). With regard to leisure activities (cultural events, creative hobbies and volunteering) respondents keeping diaries were more active than survey dropouts, with the exception of visits to the cinema and attendance at

2 Drago et al. (1999) analysed the 'busy bias' among school teachers. They found that busy teachers were the least likely to complete a time budget diary. Their results were based on a sample of 58 teachers from a high-stress, a low-stress and an average school, and might be difficult to generalize to a random survey among the general population.

pop, rock and jazz concerts. Middle-aged persons with a higher educational level responded best of all. The author concludes that those who feel a sense of duty, and those used to keeping records, respond best to a time budget survey.

Väisänen's (2002) nonresponse analysis of the Finnish time use survey showed that nonresponse rates were higher among pensioners and the unemployed, and lower among the employed and homemakers. Those who engaged in voluntary work, both formal and informal, were overrepresented among respondents, and respondents more often had a PC at home than nonrespondents (49% compared to 38%). PC-owning nonrespondents used the PC more often than respondents, however. Respondents used the computer more often for word processing and spreadsheets, nonrespondents more for playing games, chatting, e-mailing, programming and computer graphics.

Zuzanek (1999) concludes from the contributions to the round table that evidence either points to busy people being overrepresented (the Dutch case) or to an underrepresentation of 'respondents 'at the margins' of the workload continuum'. He hesitantly agrees with 'Knulst and Van den Broek in that the effects of response attrition at both ends of the social and workload spectrum may 'cancel each other out, producing a sample that is more or less representative of the population.' Examining busyness and incorporating age, type of daily activity (Väisänen), and something that might be called form literacy (Pääkönen), one might conclude that workload in itself does not seem to be an important factor in time use survey participation, but that it seems likely that different groups (not particularly the most busy ones) may not cooperate for different reasons.

Drawing final conclusions from the studies on time constraints presented here is difficult as it is not always clear which part of nonresponse is due to non-contact and which part to noncooperation. An exception is a recent study on unemployment duration (Van den Berg, Lindeboom and Dolton, 2004) which seems to confirm the manifold effects of discretionary time. They concluded that nonresponse biases the unemployment duration distribution and found both a selection effect most likely due to survey cooperation, for instance as socially isolated persons may dislike a survey interview and an interview with a case worker) and a causal effect most likely related to contactability, which occurs when a persons who has recently found a job may be more often away from home.

3.3.6 Socio-structural correlates of nonresponse: a retrospective

In discussing low response rates, 'everybody' knows which groups have little inclination to cooperate; big-city inhabitants, apartment-dwellers, ethnic minorities, elderly persons and singles are all difficult. In more in depth-discussions and studying the literature (especially the inventories of Goyder (1987) and Groves and Couper (1998), it turns out that relating socio-structural characteristics of households and individuals to response rates is far from simple. Firstly, many studies do not distin-

guish between noncontacts and refusals; this may conceal the fact that the elderly may be easier to contact, but less likely to cooperate. Secondly, the fieldwork design may partly determine who is likely to not respond. If calls are made in the evenings, for instance, employed persons may be well-represented and elderly persons who are afraid to open the door to strangers may be underrepresented. If calls are made during working hours, the employed will be underrepresented and elderly persons overrepresented. Thirdly, different recruitment or interview modes may also work out differently for different groups. Ethnic minorities may be more difficult to contact as they more often have an unlisted phone number and elderly persons who are afraid to let someone into their house may be less reticent in granting a telephone interview. Fourthly, socio-structural characteristics may need to be refined or may be interrelated which makes it difficult to isolate which factor determines survey cooperation (for instance, elderly persons are not active on the labour market, apartments are more common in big cities and ethnic minorities often live in big cities). Fifthly, single indicators may confound different aspects. Being single may mean being a wealthy, outgoing yuppie, or being a poor, elderly widow. Living in a multi-unit may mean living in a run-down flat in a poor neighbourhood or living in a luxury apartment building (with a uniformed doorman and room service) in a sought-after part of an attractive city. Just identifying single variables as determinants of nonresponse may be way off the mark and may not be helpful in either enhancing response rates or correcting for bias.

Counter-examples may help in analysing why response rates among specific groups are lower. Ethnic minorities used to have higher response rates than the original inhabitants of the Netherlands, aged people in senior housing facilities are eager to be interviewed and invite the interviewers to come back for other interviews or regular visits, fieldwork agencies do not have so many problems in the big city where their business is set up, and students (young, urban singles) are impossible to catch at home but quite compliant when it comes to answering questionnaires. This suggests – as most nonresponse researchers know – that socio-structural characteristics are not the determinants of nonresponse, but indicators for underlying social and psychological constructs explaining the likelihood of nonresponse dependent on the setting and mode of interview, the topic of the questionnaire and possibly cultural differences. As Groves, Cialdini and Couper (1992, p. 478) observe when they discuss characteristics of the sample person as factors in response behaviour: 'We do not believe these factors are causal to the participation decision. Instead, they tend to produce a set of psychological disposition that affect the decision. They also affect the initial approach of the interviewer to the sample unit.' Consequently, giving an overview of individual and household correlates of nonresponse may not advance knowledge on nonresponse, as the results so far indicate that background variables do not explain response behaviour, and do not help answer questions such as: 'Is the low response rate of ethnic minorities a consequence of urbanicity, income or other social or cultural factors?' or 'Why were the response rates among

these groups initially so high, when the socio-economic position of this group was definitely not much better than it is now?' The next chapter will focus on the causes of nonresponse.

4 Push and pull factors in survey cooperation

4.1 Beyond background variables

Chapter 3 identified the socio-structural variables that often correlate with nonresponse. This chapter will go beyond these correlates and firstly identify (section 4.2) social isolation and social involvement as a possible determinant of refusal and cooperation will also be discussed here; this section also investigates whether nonrespondents perhaps simply do not like surveys or 'this particular survey'. It might seem that nonresponse researchers must be becoming rather desperate to revert to suggesting that the major cause of survey cooperation is liking surveys or being interested in the topic. Research findings in this area, however, provide an insight into both the process of survey cooperation and nonresponse bias. Section 4.3 discusses the costs and benefits of cooperating in a survey, and social exchange theory as a framework for weighing costs and benefits. The section ends with a description of tailoring and maintaining interaction as interviewer strategies to enhance response rates. Section 4.4 looks back at the different approaches and goes into the ethics of survey research and response enhancement.

4.2 Buying surveys: why people cooperate

4.2.1 Social isolation and social involvement

Rather than looking at different socio-structural characteristics, many authors point to a single mechanism that might lie behind the lower cooperation of specific groups, viz. social isolation or its opposite, social involvement. Social isolation (Groves and Couper, 131) may manifest itself in the absence of shared norms, or the lack of guidance by norms of the dominant culture. Lower participation in official surveys, or surveys being seen as coming from vested organizations, could be expected from those who score low on civic duty, who are cynical about political institutions, and who lack trust in governmental organizations. There are both socio-structural and social psychological aspects to social isolation. If elderly people, single-person households, ethnic minorities, mobile households, people out of employment and big city dwellers are more likely to refuse, this might be due to the fact that they occupy a more isolated or less integrated position in society. Groves and Couper (1998, 152) analysed these background variables as indicators of social isolation, and – after taking correlating factors into account, such as high-cost housing in the US being in urban, high density areas, and the tendency for the elderly to live alone – concluded that: 'Multiple indicators of social integration (or isolation) share strong marginal effects on cooperation. Households with young children or young adults tend to cooperate, single-adult households tend not to cooperate.'

Social isolation may be indicated by the socio-structural variables mentioned above, but may also be related to the socio-psychological make-up of individuals. Studies in which a direct relationship can be shown between survey participation and social isolation, social exclusion and lack of social involvement, can shed more light on this. De Kruijk and Hermans (1998) found that law-abiding citizens that were more likely to vote were overrepresented in a survey. Loosveldt and Carton (2001) used direct measures of social involvement in explaining refusal in the second wave of a panel survey. Persons who scored highly on utilitarian individualism in the first wave, which means that they were more interested in personal gain and less in societal well-being, were more likely to refuse participation. In a study by Loosveldt and Storms (2003) political cynicism was negatively related to survey participation, and social involvement and trust in political organizations positively.

Political interest and knowledge may make a survey more interesting and thus increase cooperation (see section 4.3). Lack of political interest and knowledge may point to an isolated position in society in general, and thus decrease cooperation. According to Brehm (1993, p. 64) '... people who are most likely to participate in polls are also more likely to participate in society generally'. The relationship between political interest, trust in political institutions and survey participation is a central issue in Dutch studies on nonresponse in election surveys (Visscher, 1995; Smeets, 1995; Voogt, Saris and Niemöller, 1998; Aarts and Van der Kolk, 1999; Van der Kolk, 2001; Dekker, 2002; Van Goor, 2003; Voogt, 2004). Aarts, Denters and Smeets (1998) hypothesized that nonresponse to election surveys was due to lack of political interest but could not confirm this. Voogt (2004) presents a comprehensive overview of nonresponse in election studies, promotes a mixed-mode approach, presents ways to obtain information on nonrespondents and gives firm empirical evidence on the relationship between electoral participation and both social involvement and survey participation. To correct for nonresponse he recommends always asking each nonrespondent whether he or she voted at the last national election.[1] Other studies indicate that voting behaviour might be an incomplete indicator of both social involvement and survey participation, because non-voters might be more diverse than generally assumed. Dekker (2002, p. 137) for instance distinguishes four types of non-voters in the Netherlands in the nineties: elderly political cynics (low political involvement, not very active, dissatisfied but not likely to protest, no higher education), younger political cynics (similar to the elder ones, more likely to protest), the politically unconcerned (less cynical about politics than the former groups, somewhat higher education level) and the politically involved (the smallest group of non-voters, politically involved and active, equally likely as the younger cynics to protest, more higher educated people).

1 A discussion of this 'central question' in an international setting immediately showed the problems of doing this due to diverse electoral systems and to voting being mandatory in some countries.

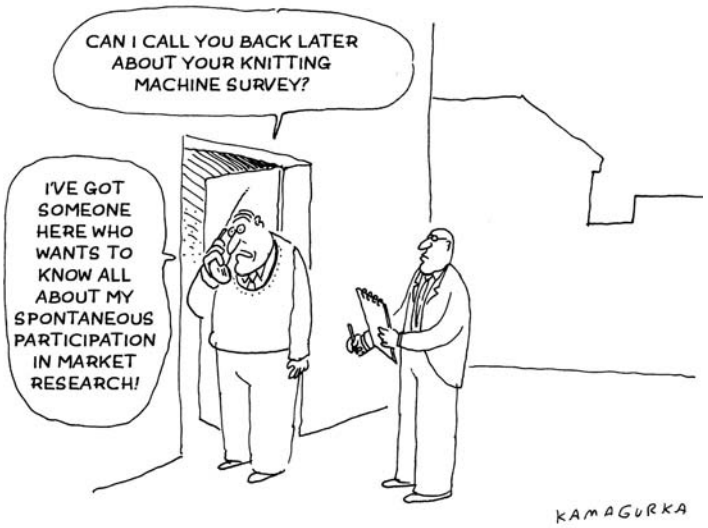
The cynics from the first two groups (together almost 60%) are stereotypical of the dissatisfied and passive non-voter (and presumably nonrespondent). Although voting behaviour might not be relevant cross-nationally and does not characterize a homogeneous group, it seems promising to try to find how voting is related to nonresponse bias. This recommendation receives support from the results from Groves, Presser and Dipko (2004) who found that political contributors showed a relatively high cooperation rate in surveys independent of the topic of the survey.

Social isolation came up in the studies of the amount of discretionary time. Participating in a survey might deplete the very scarce (time) resources of someone who is very busy, and thus might decrease the willingness to cooperate in a survey. On the other hand, people who are busy might be more involved in society, and thus more inclined to participate. The concept of social isolation, and the link to the norms of the dominant culture, might explain the results of Pääkkönen (1999) that respondents spend more time on leisure activities (cultural events, creative hobbies and volunteering) with one exception where refusals were more active, namely visits to the cinema and attendance at pop, rock and jazz concerts. Leisure activities in general might be part of the dominant culture, whereas cinema, pop, rock and jazz concerts are closer to a youth subculture less inclined to show law-abiding survey participation behaviour. A similar observation could be made with respect to Väisänen's (2002) analysis of time budget data. Respondents more often engaged in voluntary work and more often had a PC at home than nonrespondents. PC-owning nonrespondents used the PC more often than respondents; though. PC-ownership may be related to wealth, PC use may be related to form literacy and cognitive skills, and the type of PC use (word processing and spreadsheets for respondents and playing games, chatting, e-mailing, programming and computer graphics for nonrespondents) might indicate the difference between the dominant culture of serious citizens working or doing their bookkeeping at home, as opposed to a subculture of 'cyberkids'. This suggests that survey participation may be firmly embedded in the socio-cultural mainstream.

4.2.2 *Reasons for refusal and attitudes towards surveys*

In the previous section the central question was 'Who is likely to respond?'. Here, the focus shifts to the question 'Why do people participate and is there such a thing as an attitude towards survey participation?' Nonresponse researchers have addressed these questions by measuring the general attitude towards surveys, the attitude towards particular surveys and the impact of substantive survey characteristics such as topic and sponsor. They did this by recording doorstep interactions on tape or paper forms (Couper, 1997; Campanelli, Sturgis and Purdon, 1997; Loosveldt and Storms, 2001), recording reasons for refusal, incorporating questions on surveys in surveys (Van Tulder, 1977 and 1978, Singer, Van Hoewyk and Maher, 2003; Loosveldt and Storms, 2004), conducting surveys on surveys (Goyder, 1986; Werkgroep dataverzameling, 1996; Stocké and Langfeldt, 2004) and by mounting follow-up surveys

among respondents whose attitude towards a survey is known from an earlier survey (Hox, De Leeuw and Vorst, 1995; Rogelberg et al., 2003).

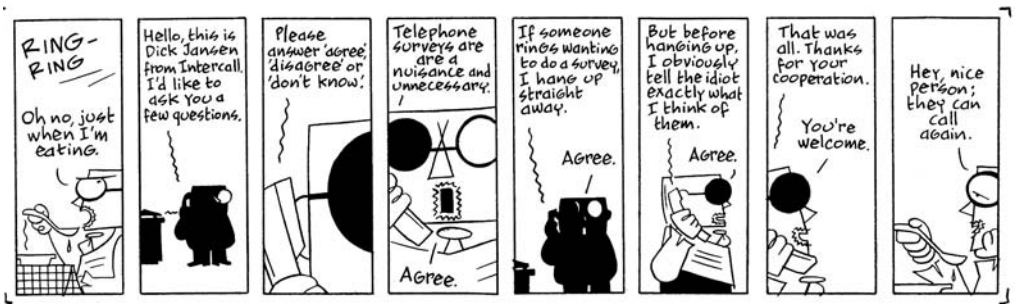


One might wonder if recording reasons for refusal, analysing the answers on opinions on surveys from the survey's respondents, or mounting surveys on surveys, are feasible ways of measuring survey attitudes. The reasons expressed for refusal may not reflect the sample persons 'real' meaning and just be a way of getting rid of the interviewer as soon as possible and avoiding discussions. Smith (1984, 486) remarked: 'Review of detailed interviewer comments on final refusals indicates that reasons for refusing are not always clear and that some people seem to offer a string of reasons hoping that the interviewer will expect one and leave them alone, and Brehm (1993, p. 51): Skeptical readers might wonder whether the refusals' reasons are just the first coherent thought that pops into mind, and not the cause behind the decision. By this thinking, the refusal might feel threatened by the interview, but tell the interviewer, 'I'm too busy.' Other sceptics might believe a refusals' reason is just what the person thinks will convince the interviewer. More extreme sceptics might even doubt that the refusal is capable of knowing why he or she chose to refuse. Such skepticism is probably well-founded. On the other hand, why not treat what people say as useful information?' Despite these doubts, this section will present empirical evidence on the relationship between reasons for (initial refusal), personality traits, survey outcomes and persistence of refusal.

Smith (1984, p. 481-485) studied the propitiousness and inclination to participate in a survey of temporary refusals. Propitiousness is a situational factor, determining whether the interview request is convenient at that particular occasion when the respondent has other activities (going out, napping, having a headache, having dinner, entertaining visitors). Coming back at a more suitable time or making an appoint-

ment to do this will be a useful strategy. Independent of the level of propitiousness, respondents might differ in their inclination or willingness to be interviewed. Smith categorized some causes of being unwilling to transient problems related to the ones mentioned above (having family problems or work-related pressures). Other types of unwillingness, such as fear of intrusion, unpleasant experiences with earlier surveys, and general concerns about confidentiality and privacy, might be more difficult to overcome. Inclination may also be determined by personality traits, such as suspiciousness, misanthropy, misogyny, reclusion and paranoia. Smith feels that the multiple character of refusals might be a reason why he did not find much difference between cooperative and temporary respondents, because converting refusals may have worked only for refusals due to situational factors or possible transient inclinational factors, and not for permanent types of disinclination.

Sigmund-Peter de Wit



Couper (1997) related reasons for refusal, the statements made by respondents in the introductory conversations and their answers in the subsequent interviews. He concluded that those who show a lack of interest in the topic are less likely to grant an interview and, if they do cooperate, produce less meaningful data (more items missing) and differ in the substantive responses from those who do not express a lack of interest. Those who refuse to cooperate because they are 'too busy' show little systematic variation, which might mean that busyness is just a polite (and easy) way of saying 'no' or that busyness is not related to the topic of the survey. Couper's results were confirmed by Campanelli, Sturgis and Purdon (1997, pp. 4.21-4.24). Consequently, the refusal of those who are not interested in the topic of the survey may lead to biased results. The same concern comes up again and again in Dutch studies on nonresponse in election surveys (see section 4.2). Laurie, Smith, and Scott (1999, p. 279) found that an initial refusal was more likely to be final when the reason for refusal was survey-related (confidentiality, too long, waste of time) rather than respondent related (too busy, not able, stressful situation, refusal by proxy). Voogt (2004, pp. 41-45) studied the difference between nonrespondents who refused because they claimed not to be interested in the survey and those who refused for other reasons. There was no relationship between reason for refusal and willingness

to participate in a short telephone interview (52% versus 47%). Both groups were fairly similar, except that the 'not interested' refusers more often agreed with the statement that politics are too complicated for them.

Besides recording reasons for refusal or introductory respondent reactions, information on survey attitudes can be obtained by asking questions on surveys in a regular survey or mounting a specific survey on surveys. Survey attitudes of survey participants might be expected to be rather positive, if only to reduce cognitive dissonance. This bias can be prevented by asking questions on surveys to a group that for one reason or another shows complete response, and study their future response behaviour. One might contest that surveys on surveys are a paradoxical way of obtaining information as employing an instrument to measure its own performance is immediately contradictory (see Goyder, 1986). Surveys on surveys are especially useful when external sources of evidence can be used, such as information from the sampling frame, or when survey attitudes can be linked to future response behaviour. Another reason why surveys on surveys might yield useful information is because, as mentioned above, involvement in the topic seems to enhance response propensity. Consequently (see Goyder, 1986, p. 39) surveys on surveys may attract both opponents of surveys and professional respondents, and may be saved from terminal selection bias. Schnell (1997, p. 194) remarks on the interesting fact that part of the participants in a survey agreed that surveys are an utter waste of time, and in a Dutch survey on surveys (Werkgroep dataverzameling, 1996), a number of respondents indicated that they never cooperated in surveys.

Hox, De Leeuw and Vorst (1995) analysed the response behaviour of students who had earlier filled in a questionnaire in which their attitudes towards surveys had been measured. Their study was based on the input of Robert Cialdini in the first meetings of the International Workshop on Household Survey Nonresponse, and led to the unexpected result that neither a general attitude towards surveys, nor a specific intention to cooperate in a survey similar to the one that was later presented, predicted final response behaviour very well. In a study by Rogelberg et al. (2001) willingness to participate in a future survey among students depended on two attitudes towards surveys, viz. survey value and survey enjoyment, the latter factor being the more important. Survey enjoyment was measured by three questions 'I do not like filling out surveys', 'Surveys are fun to full out' and 'I enjoy filling out surveys', and survey value by 'A lot can be learned from information gathered from surveys', 'Nothing good comes from complementing a survey' and 'Surveys are useful ways to gather information'. In 2003 Rogelberg et al. reported on a two-wave survey among students. Based on the results of the first wave, they could distinguish between active and passive nonrespondents in the follow-up survey. Failure to respond by the former, more or less accidental nonrespondent was not based on a conscious and overt *a priori* decision. The active nonrespondent, in their model, had made a conscious decision to not respond to the survey as soon as the survey is received.

Response enhancement techniques may only work for passive nonrespondents, not for active ones. Rogelberg and his colleagues found that passive nonrespondents are very similar to respondents. Active nonrespondents, however, were less satisfied with the survey sponsor, less conscientious, more likely to leave the university and less agreeable.

Singer, Van Hoewyk and Maher (1998) included questions on attitudes towards 'surveys like this' in the repeated measurement Survey of Consumer Attitudes. Answers might of course bear on the topic or on more general characteristics of the survey. Over time, more respondents felt that they should be paid (and a higher amount) and that everybody has a responsibility for doing surveys like this. They also found that respondents who said they would not do the survey again did consider it a waste of time or not useful, and that respondents who disagreed that responsible persons should cooperate, provided data of poorer quality (more don't knows, and more refusals to answer a question). Loosveldt and Storms (2001) discussed the relevance of respondents' attitudes towards surveys, gave an overview of the literature and included the Singer, Van Hoewyk and Maher items in a survey on cultural shift in Flanders. In this study the attitude to 'surveys like this' was more positive, which may be due to the topic, the interview mode (telephone in the US, face-to-face in Flanders) or to cultural factors. They also included two other general evaluative questions, on how meaningful surveys are and about the credibility of survey results. Three-quarters of the respondents found surveys (very) meaningful and less than 5% survey results (almost) never credible. A positive attitude towards surveys was related to positive experiences with surveys in the past, and with willingness to cooperate in the future. The authors compared survey attitudes as measured at the end of the interview with doorstep reactions, and found that – as could be expected – positive reactions are related to more positive attitudes, and negative reactions to more negative attitudes, but also that asking questions about the survey on the doorstep is indicative of a negative attitude. Finally, their results suggest a positive relationship between attitudes towards surveys and both trust in the working of democratic institutions and the attitude towards unpaid work.

Further research from Loosveldt and Storms (2003) generally confirms their earlier results, with one surprising difference: the percentage of respondents that did not want to cooperate in a future survey had almost doubled (from 9.8% in 2000 to 17.9% in 2002). This might be due to the longer interview duration in 2002 (74 minutes in 2000 and 91 minutes in 2002) and the greater perceived difficulty of the questions. The authors (pp. 359-363) report lower response rates among elderly women and persons with a lower education level. This could be due to their more negative attitude towards surveys. However, sex differences in attitudes were not significant. Persons with a lower education were more positive towards surveys than those with a higher education background, and the elderly more than younger persons. They did find a relationship with discretionary time (see section 3.3.5), and that those in paid

employment and those with children younger than 18 had a more negative attitude towards surveys than others. In a multivariate analysis only the (weak) effect of age and paid work remained, and the effect of age disappeared after taking into account the evaluation by the respondents of the general atmosphere during the interview.

Two conclusions can be drawn from the above results . The first is that meaningful information can be obtained by studying reasons for refusal, doorstep comments of both respondents and nonrespondents, the answers to questions on survey participation in regular surveys, mounting surveys on surveys and studying surveys in an experimental setting. The second conclusion is that all studies presented above, with the exception of Hox, De Leeuw and Vorst' 1995 study, seem to indicate that survey noncooperation might be more or less 'at random' when situational factors are the reason for nonparticipation. These nonrespondents are busy, have minor household crises to deal with, simply do not feel like doing it and have no strong feelings on the value and enjoyment of surveys. They might well participate if the interviewer were to come back later or if they could have completed the interview at their own convenience. However, if nonparticipation is largely determined by the topic or the sponsor of the survey, nonresponse will be 'not at random', and cannot be ignored. This substantive aversion should be compensated for by external incentives. In addition, persons who harbour a strong dislike of surveys will be more difficult to convert than persons who do not cooperate for accidental reasons. This does not necessarily lead to bias, except when survey attitude is related to survey topics, such as trust in government.

The assumption that the attitude towards a survey is largely determined by the topic of the survey and the type of sponsor is confirmed by many studies (Snijkers, Hox and De Leeuw, 1999, p. 191). Government and academic surveys receive higher response rates, probably because they appeal more to a sense of civic duty and authority than commercial surveys (Dillman, 2000, p. 20; Groves and Couper, 1998, p. 139). In designing advance letters, it is considered good practice to have these signed by important public figures (a minister, a professor, the mayor). De Kruijk and Hermans (1998) found a substantial difference in response rates in a telephone survey, depending on whether the city council of Amsterdam (66%) or the market research organization (54%) was mentioned. The difference in response rates had no effect, however, on the outcomes of the survey. An exception is the recent study from Stocké and Langfeldt (2004) in which the attitude towards surveys did not depend on whether the respondent's last survey interview was conducted either by a public authority or scientific organization or by a commercial market research firm.
In introducing surveys, it is considered good practice to emphasize that the topic of the survey is relevant to the interviewee, assuming that topic saliency is related to response behaviour (Groves and Couper, 1998, pp. 286-288). In the Dutch survey on surveys mentioned above (Werkgroep dataverzameling, 1996) respondents favoured topics such as health, employment, education, social security, housing, hobbies and leisure, environmental issues, traffic and discrimination, and disliked surveys

on income, politics and consumer behaviour. The main reason respondents in this survey gave for noncooperation was that they did not like the topic. Brehm (1993, pp. 61-64) shows that interest and informedness affect the respondents' relationship to surveys on political issues, and the respondents' chances of complying. Voogt, Saris and Niemöller (1998) found that the willingness to participate in a political survey was much higher among those who were interested in politics. In a study from Groves, Presser and Dipko (2004) persons cooperated at higher rates in surveys on topics of likely interest to them. Political contributors cooperated with all the topics, however, which might point more to social involvement than to topic relevance as a determinant of survey participation.

Te Riele (2002) presented an inventory of topic-related response behaviour in surveys of Statistics Netherlands. Although she does not always distinguish between non-contact and noncooperation, her results are impressive: people who are less mobile are less likely to participate in the mobility survey (a result also found by Mathiowetz, Couper and Butler, 2000), converted refusals in the housing demand survey are less likely to want to move, persons with health problems are overrepresented in the health survey (a result which contradicts the findings from Cohen an Duffy (2000), see section 3.2), persons who are interested in politics are overrepresented in the national election studies, and – in general – there seems to be a relationship between social participation and survey participation.

The results of this section are twofold. Firstly, there is evidence that social involvement, political interest and being part of the cultural mainstream enhances the probability of survey cooperation, regardless of the topic. Secondly, interest in or relevance of the topic is also likely to be positively correlated with survey cooperation. Finally, accidental refusals may be more easily converted than refusals related to surveys, the topic, the research agency or the sponsor. The latter type of refusal is more likely to cause bias. This leads to the question of how survey cooperation can be made attractive independent the particular character of a survey. This will be discussed in the next section.

4.3 Selling surveys: strategies for response enhancement

4.3.1 Cost and benefits of being interviewed

Groves and Couper (1998, pp. 121-125) state that according to the rational choice theory of decision-making an individual would weigh the costs of participation in a survey against the perceived benefits, the cognitive burden and confidentiality risks. There is not much evidence for a fully rational approach, however, not with regard to survey participation, nor in other areas. Contrary to the rational choice approach, for instance, is the observation that the decision to cooperate can be taken within seconds (Sturgis and Campanelli, 1998, p. 7). Dijkstra and Smit (2002) found that if the respondents answered 'good evening' to the first introduction in a telephone

interview, this was highly indicative of future cooperation. Consequently, survey researchers have abandoned the strictly rational choice approach and moved on to theories on opportunity costs, social exchange and social isolation. Social isolation and the opportunity cost approach were elaborated in section 3.3.5. This section will focus on the social exchange framework, in which it is assumed that participating in a survey has certain disadvantages and certain benefits that need not be strictly material by may also be social (Brehm, 1993, pp. 59-68, Dillman, 2000, pp. 14-21). The net outcome should determine survey participation.

The following issues most frequently come up in overviews of survey costs (Groves and Couper, 1998; Dillman, 2000): interview length, frequency of survey requests, misuse of surveys as a sales pitch, the cognitive burden of the interview, possible embarrassment and inconvenience caused by survey cooperation, an invasion of privacy, and being treated irreverently. The first factor related to the costs of being interviewed is interview length the effect of which has already been discussed in section 3.3.5. The costs of participating in a survey would include time required to complete the interview and the lost opportunity to perform other activities. Although there does not appear to be a clear-cut relationship between interview duration and survey cooperation, there seems to be a universal tendency to downplay the length of an interview in the survey introduction. The frequency of survey requests may result in survey fatigue, or survey saturation. According to Goyder (1986) '... the more times a person reports requests for survey cooperation, the more unfavorable is his or her attitude towards the method'. The frequency of survey requests may also counteract scarcity arguments used by the interviewer in the recruitment phase (see below). Selling under the guise of a survey (sugging, see Groves and Couper, 1989, p. 172) or using surveys as a sales pitch is something more and more people are becoming wary about. De Leeuw (2001) even suggested starting every telephone survey introduction with 'I am not selling anything'. In her experiments this led to a small but positive effect on response rates. Stocké and Langfeldt (2004), however, found that the misuse of survey interviews as a door opener to sell goods and services had no effect on respondents evaluation of surveys in general. The cognitive burden of the survey, sensitive or difficult questions (Tourangeau and Smith, 1996), a perceived invasion of privacy (Singer, Mathiowetz and Couper, 1993; Singer, Van Hoewyk and Neugebauer, 2003) or the feeling of being treated disrespectful (Morton-Williams, 1993, p. 105; Dillman, 2000, pp. 15-16) may also be a final reason for refusing cooperation.

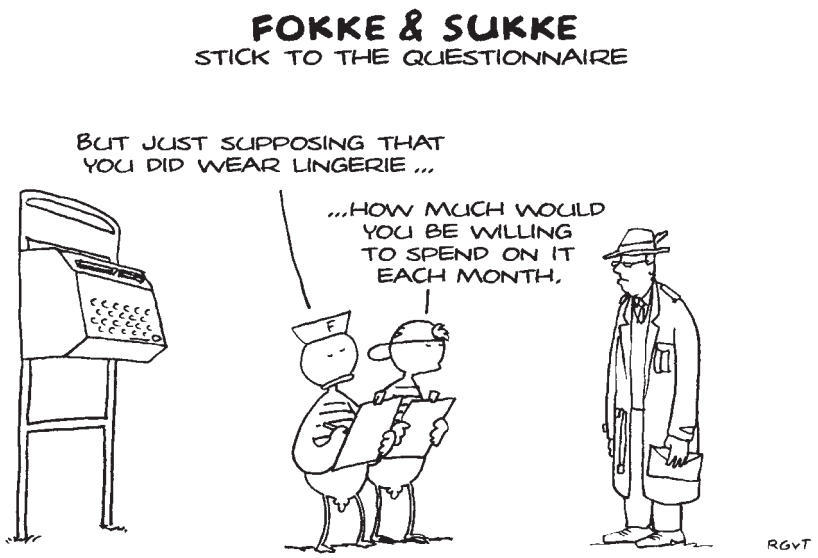
The benefits include avoiding even less attractive activities, the satisfaction of being part of a socially useful or academically interesting enterprise, the enjoyment of thinking about interesting topics, interaction with the interviewer, the gratification that one's opinion was sought, the satisfaction of fulfilling a civic duty (see section 1.3 for the Dutch attitude to civic duty as a reason for survey cooperation, and the previous section for a review of the importance of sponsor and topic), receiving social validation, being respected and valued, being selected as one of many and representing other

people, having the chance to influence government policy making, obtaining information in advance and feedback on the outcomes, and finally receiving a tangible reward or monetary incentive. Survey benefits are to some extent the mirror image of survey costs. Whereas an overload of survey requests may lead to survey fatigue and nonresponse, the perception of being selected as one of many and representing other people (scarcity), and even being able to influence government policies may induce a respondent to cooperate. Whereas answering a long and boring questionnaire may take precious time that could otherwise be spent on more fulfilling pastimes (e.g. watching TV), discussing an interesting subject with an interested interviewer might be a highly rewarding activity. These survey benefits are reflected in arguments that interviewers use in persuading respondents to cooperate.

Twenty-five years ago, survey respondents were asked to score their initial reaction to the request to participate in one of the following categories (Van Tulder, 1977, 1978)

a To tell the truth, I wanted to refuse, but I cooperated after all
b I was somewhat hesitant, and considered for a while not to cooperate.
c I did not exactly like it, but to consider a refusal did not cross my mind.
d I thought something like 'Why not, if I can do someone a favour'.
e I enjoyed hearing what kind of questions is asked.
f In the newspapers, on television, on the radio, you always hear about people being part of surveys and now they ask my opinion. I liked that.

Most of these answer categories are related to different social exchange aspects (civic duty, altruism, substantive interest, scarcity). Sadly, these arguments were seen as an ordinal scale and respondents could select one alternative only. Most respondents scored the more positive categories (e, f and especially d). After computing average scale values, it turned out that, the average in Amsterdam was much lower than everywhere else. They also found that younger people had a higher score (and were less likely to 'almost' refuse) than older people, and also that lower and higher SES groups might be more prone to refuse cooperation.



Hox and De Leeuw (2002) studied, among other things, the use of scarcity arguments (*this* is the chance to give your opinion) and social validation arguments (most people participate, emphasize positive aspects of participation) across countries and found that neither argument was often used in the Netherlands, which might partly explain the low response rates in this country. They also suggest, however, that social validation arguments might remind people too much of a sales pitch and thus might even be counterproductive. Loosveldt and Carton (2001, 2002) observe a negative relationship between utilitarian individualism, which means being less prone to do things of general interest and less altruistic, and the attitude towards surveys and likeliness to cooperate in subsequent waves of a panel survey. This indicates that the benefits of partaking in a survey are higher for those who value maintaining a good relationship with their fellow citizens, and lower for those who strive for personal (material) interest and success. Utilitarian individualism is seen as an indicator of potential social integration and a reduced solidarity between people, and is therefore closely related to the concepts of social involvement and social isolation mentioned above.

Other rewards of survey cooperation are obtaining information in advance and feedback on the outcomes, and finally obtaining a tangible reward or monetary incentive. The possible benefits obtained from incentives and the efficacy of prior notification will be outlined below, preceding a discussion of more holistic techniques of obtaining survey cooperation.

4.3.1 *Incentives*

David de Vaus, the editor of the series of Social Surveys (see section 1.2) gives on his website on social research users.bigpond.net.au/david-devaus/Reading-guides/using_incentives.htm a useful list of studies on the use of incentives. Here only a small number will be mentioned. Stoop and Oudhof (1988) studied the effects of incentives in an experimental study on time use. Incentives were mainly effective in preventing dropout out in burdensome variants rather than in persuading respondents to cooperate. Willimack and her colleagues (1995) found that enclosing a pen with the advance letter increased response rates in an indirect way. The pen did not increase the value of the numerator (by making more persons respond) but decreased the value of the denominator, as more ineligible sample units were identified, probably because with a pen letters were larger and required more postage and were consequently more often returned to sender (a similar effect occurred in another survey when questionnaires were sent by classified rather than regular mail). In their study, incentives were effective in the initial phases of trying to obtain survey cooperation. Warriner et al. (1996) compared cash, lottery and charitable donation incentives and found that a cash incentive in between $2 and $5 was most effective. They related other social exchange aspects to the type of incentive and suggested that 1) 'the kind of respondent who perceives legitimacy to survey questions about a salient topic such as environmentalism, reinforced with high-credibility sponsorship, perhaps is unmoved by lottery incentives', and 2) 'if genuine altruism is a motive for survey response, the promise of a charitable donation on behalf of the respondents is a clumsy trigger'. Martin, Abreu and Winters (2001) did not find any evidence that incentives have unintended effects on intrinsic motives for survey participation, such as altruism, civic duty, or interest. Neither did they find evidence that larger incentives (up to $20 or $40) worked better (or worse) nor that the level of the incentive interacted with the financial situation of the household. Couper, Ryu and Marans (2002) tried to customize the type of incentive to the topic of the surveys (visiting parks) without much success. Davern et al. (2003) found no effect of monetary incentives on survey quality in a survey comprising mostly factual questions. Nicolaas and Stratford (2004) studied the effect of no incentive, $5 and $10 incentive per household member in a survey in which every household member had to cooperate. Both incentives lead to a higher response rate, but the added value of $10 was only marginal. The larger incentive lead to an overrepresentation of lone parents with dependent children, who obviously could use the money. Prepaid incentives are generally considered to be most effective, although handing out prepaid incentives is not always feasible. Nicolaas and Stratford (2004) summarize the literature as follows:

> 'The theory that is often used to explain why prepaid incentives are more effective than promised incentives is the 'norm of reciprocity': people feel obligated to respond to positive behaviour with positive behaviour in return. If we give a respondent an unconditional incentive, they will feel obligated to reciprocate by

responding positively to the survey request. Prepaid incentives are more effective than promised incentives because obligation is a stronger motivational factor than inducement or 'bribery'.'

Eleanor Singer and her colleagues have extensively studied the effects of using incentives. Singer, Van Hoewyk and Maher (1998) found that incentives seemed to reduce item nonresponse (in an opinion survey) and that people who had received a monetary incentive in the past were more likely to expect a payment in the future, although this expectation does not seem to affect actual survey cooperation. Singer, Groves and Corning (1999) discovered that three-quarters of their respondents considered it unfair that those who initially refused cooperation in a survey were offered a reward, whereas those who immediately cooperated got nothing. This judgement had no influence, however, on there willingness to participate in a future survey run by the same organization that used these unfair procedures, nor in their actual willingness to participate in a later survey from an ostensibly different survey organization. They also discovered that enclosing a $5 incentive in a mail survey had a large effect on participation. Singer et al. (1999) give a valuable overview of the literature and found again that a moderate cash reward seems to universally beneficial to survey quality, even when the survey burden was low. In 2000, Groves, Singer and Corning presented their leverage-saliency theory on survey participation in which they combine a social exchange approach with tailoring doorstep behaviour to respondent concerns (see below) and study the relationship between civic duty as an internal and cash rewards as an external incentive to survey cooperation. The positive effect of a (monetary) incentive appears to be diminished when community involvement is a likely motivation for cooperation. In a sequel, Groves, Presser and Dipko (2004) could not confirm that a financial incentive leads to higher cooperation from those who attach higher leverage to personal benefits, however.

The evidence from the many studies on incentives is not entirely conclusive. Most researchers seem to agree that a small incentive is generally seen as a nice gesture, both by interviewers and respondents, that a large incentive is not necessarily more effective than a small one, and that incentives do not have a detrimental effect on survey quality.

4.3.2 *Prenotification letters and phone calls*

It is generally assumed that the decision to cooperate in face-to-face surveys takes place on the doorstep, during the interaction between interviewer and householder. If an advance letter has been sent, the decision process may have started even earlier, at the reception and perusal of this letter. Advance letters are nowadays considered to be an important means of conveying information on the survey and inspiring trust in the fieldwork organization and the sponsor, but in the earlier literature on nonresponse they were seen as a dangerous procedure. Gorden (1980, p. 253) argues strongly against using advance letters and suggests that it would be best to introduce

a survey through the news media. Morton-Williams (1993, p. 61) explains how interviewers used to be dubious about the value of advance letters, but have come to see them as useful. She found that they increased response rates, and mentions as a practical problem that in address samples, advance letters are generally addressed as to 'The Householder' or to 'The Resident' which may reduce trust or the pleasant feeling for the prospective respondents of being specially selected . Dillman et al. (2002, p. 11) considers an advance letter as a good thing, or at least accepted practice in survey research. Design features that may be salient to the respondent can be highlighted in an advance letter. They also mention that advance letters increase the interviewers' confidence while seeking cooperation (see also section 4.3.5). There appears to be no empirical evidence that an advance letter might lead to lower cooperation because the sample person will be prepared and better able to resist the interviewer's request. This might be because respondents do not read advance letters anyway. 'The measure of attempts for the list sample excludes the initial mailing, which offered respondents a chance to opt out of the survey. From reports of respondents' comments during the field period, it appears that many of them were unaware of having received this mailing' (Kennickel, 2000, p. 11). Groves and Couper (1998, p. 276) seem to consider an advance letter as a matter of course. They provide a list of design features of advance letters (official stationery, personalizing name and address, providing information, etc.) that can be used to manipulate a variety of influences known to affect survey participation, and also mention that interviewers believe that advance letters enhance their own confidence in seeking the interview request at the first contact with the sample household.

A number of fieldwork organizations in the Netherlands no longer allow their interviewers make 'cold calls'. To minimize the costs of fieldwork and to motivate interviewers by having them call on cooperative sample households only, they have shifted the recruitment phase of face-to-face surveys to call centres that arrange appointments for face-to-face interviews. Other organizations are wary about preliminary telephone contact, because that would make refusal all too easy, as is shown by the lower response rates in telephone surveys. Still other organizations recommend or allow contact attempts by telephone only after a number of unsuccessful house calls. Paradoxically, while it is feared that recruitment by telephone might decrease response rates in face-to-face surveys, there is empirical evidence that telephone prenotification increases response rates in mail surveys (Dillman, Gallegos and Frey, 1976; Schlegelmich and Diamantopoulos, 1991).

There is not much empirical evidence on the pros and cons of using the telephone for recruiting respondents before making home calls for the actual interview. Gorden (1980, p. 253) is clearly not in favour of alerting households to an upcoming interview request: 'In trying to reduce the number of refusals a common mistake is made by inexperienced surveyors who first make a telephone call to set up an appointment at a convenient time for the respondent. This strategy has been repeatedly proven

ineffective by experimental studies'. He gives just one example of the possible adverse effects, referring to Brunner and Carroll (1967) '... the refusal rate among the respondents who were first contacted by phone was much higher than in the cases where the first contact was the personal call. Those contacted by telephone had a refusal rate of 63 percent in contrast to only 33 percent for the face-to-face contacts. In the sample sponsored by the University the rates were 55 percent and 16 percent, respectively'. Groves (1989, p. 209, 210) also refers to the negative outcomes from Brunner and Carroll but does not mention any other evidence against. He criticizes their study: 'There is no description provided about the nature of the prior telephone contact, but it is noteworthy that the interviewers in the survey were students in the marketing class. Groves and Magilavy (1981) attempted a prior telephone contact before a random digit dialed telephone survey ... They found negligible differences between response rates.' And then he extensively discusses the importance of advance letters. Brunner and Carroll (1967) provide no information at all on the content of the prior telephone call. Morton-Williams (1996, p. 62) found only minor increases in nonresponse, and a major decrease in fieldwork costs when first contact was made by telephone. She recommends sending an advance letter to lend authority to the study.

4.3.3 Foot-in-the-door or door-in-the-face

Advance letters and incentives can be seen as a small-time salesman approach. The ultimate salesman approach is the foot-in-the-door technique, i.e. using an initial act as self-perception leverage enabling an ensuing larger act (Goyder, 1987, p. 164, Hippler and Hippler, 1986, Dillman, 2001, p. 19, 248). Hippler and Hippler (1986, p. 27) do not think this technique is effective in face-to-face surveys, because its effectiveness seems to depend on a distance of time (a minimum of two days) between the first (minimal) and the second (larger) request, and on a change of the requesting person. Goyder definitely has mixed feelings, as he regards this technique as an abhorrent exponent of behaviour modification strategies infringing the freedom of choice of respondents, and in addition does not consider it to be very effective. Dillman (1999), also working from social exchange principles in his tailored design method, reported excellent results using the foot-in-the-door technique by a series of successive requests to the respondents, each being more 'costly' than the earlier one. Hox & De Leeuw (2002, pp. 115-116) conclude from a study of cultural differences in interviewer attitudes and behaviour that Dutch interviewers achieved extremely low scores on the attitude dimension 'persuasion orientation' and below average scores on the three behavioural dimensions in their model, viz. the use of social validation and scarcity arguments, and foot-in-the door techniques. They did not find any effect, however, of (alleged) doorstep behaviour, including foot-in-the-door techniques or consistency arguments, in their meta-analysis across countries.

The opposite of the foot-in-the-door is the door-in-the face technique (Mowen and Cialdini, 1980; Hippler and Hippler, 1986 Groves, Cialdini and Couper, 1992). This implies asking for a large favour (cooperation in a long interview) where there is

a chance of having the door slammed in one's face, and if this request if rejected, asking for a smaller favour as a concession (cooperation in a short interview).[2] Mowen and Cialdini indicate that it is important that the first request is seen as legitimate, and that the second request should be related to the first one. Ego protection mechanisms might work against the door-in-the-face technique. The authors also mention, somewhat surprised it seems, that the innocuous statement 'it would really help us out' also may increase compliance (p. 257). Hippler and Hippler's use of the door-in-the-face technique implied first asking a detailed income question and – if no answer came – asking a more general income question. This procedure was very effective and is now used in many questionnaires. As mentioned by Groves, Cialdini and Couper (1992, p. 481) the door-in the face technique might be relevant to double-sampling designs, and also for the central question technique to be discussed in the next chapter.

4.3.4 *Tailoring*

Despite the fact that the decision to cooperate in a survey may be made within the first few seconds of interaction with the interviewer, the literature presented above clearly indicates that cooperation rates can be enhanced by maximizing the advantages and minimizing the disadvantages of survey cooperation, or at least highlighting the pros and downplaying the cons. The results also indicate that individual preferences may partly determine what is a pro and what is a con, and that different aspects of the survey and the fieldwork approach may have a different impact on different respondents. Appeals to civic duty, handing out cash incentives, highlighting the intrinsic interest of a survey or referring to legitimate sponsors may not be uniformly effective. Face-to-face approaches may inspire confidence in some respondents and shy away others. Advance letters may be too complicated for people with a lower education level and too simple for the more highly educated. This suggests that flexible survey designs and the fieldwork approaches could enhance response rates.

Voogt (2004) is very much in favour of adapting the interview mode to individual preferences. The success of the Neu Kontiv design (Snijkers and Luppes, 2000; Kalfs, 2001) suggests that a respondent-driven approach in which the respondent is seen as a customer rather than a pawn would enhance response rates. Dillman (2000) found that changing the nature of appeals over successive waves in mail questionnaire surveys led to higher response rates. Continuing this line of argument, it might be ideal – not only to have respondents select their preferred interview mode – but also draft different advance letters and assign different types of interviewers to different types of respondents, and hand over incentives only to those who participate for monetary gain. This would of course tremendously complicate fieldwork logistics,

2 This technique sounds unpleasantly similar to a street robber asking you for all your money, asking for 10 Euro if you refuse and then being disappointed when you don't even go along with this fair 'compromise'

and besides the type of respondent is generally not known in advance. In its extreme form, tailoring could even be extended to other survey aspects, such as question-phrasing. This might of course increase measurement error. Schober and Conrad (1997) observe, however, that administering questionnaires in a conversational rather than a standardized way substantially reduced measurement error. Sadly, it also substantially increased the duration of the interview.

An entirely custom-made approach has its drawbacks, but a certain amount of tailoring would definitely be very useful in enhancing response rates. Experienced interviewers routinely adapt their behaviour to the perceived features and preferences of the respondent, and this can to some extent be learnt. Groves and McGonagle (2001) applied the constructs of tailoring and maintaining interaction to a training protocol design. In tailoring the interviewer emphasizes those survey characteristics that are salient to the respondent, insists when useful and withdraws when necessary. The longer interaction is maintained, the more cues the interviewer receives that can be used in tailoring, and the better the opportunity to select the most appropriate approach from the wider repertoire of persuasion techniques good interviewers have built up. In maintaining the conversation, '... the expert interviewer does not maximize the likelihood of obtaining a 'yes' answer in any given context, but minimizes the likelihood of a 'no' answer over repeated turn taking in the contact' (p. 251).

Groves, Singer and Corning (2000) and Groves, Presser and Dipko (2004) have elaborated the combination of social exchange and tailoring in their leverage-saliency theory of survey participation. According to this theory different survey attributes, such as topic, type of research organisation and presence of a cash incentive, are weighed by the respondents in their decision on whether or not to cooperate. The attributes have a different leverage to the individual respondents and can be made more or less salient to the respondent prior to the participation decision. Interviewers implicitly use this theory when they emphasize, for instance, academic sponsorship to respondents with a higher education level, and the handing out of a gift to less wealthy respondents. As mentioned above, the first test of the theory showed that the effects of incentives were smaller for those with high community involvement than for those with low community involvement. The second study, however, found no significant relationship between incentive and personal gain orientation.

Experienced, motivated interviewers might be better at tailoring, better at downplaying the costs, and better at highlighting the benefits of partaking in a survey. Pondman (1998, p. 108) concludes that '... It is beyond dispute that it is possible to motivate interviewers in such a way that nonresponse due to refusals will be reduced'. As motivating factors for interviewers she mentions the importance of providing extensive information on the content of the interview, conveying the importance of high response rates across groups, 'and to make them believe that they are really capable of persuading subjects to participate'. Houtkoop-Steenstra and Van den Bergh (2000, p. 295) studied

the relationship between motivation and tailoring and suggest that the conversational introduction to surveys may have been effective because '... the interviewers who formulated their own introduction were more motivated because they enjoyed what they did, as several told us after the experiment'. Däubler (2002, p. 42) found in the German SOEP-panel that experienced interviewers achieved higher response rates. Lemay and Durand (2002) discuss the importance of motivation and self-efficacy, although their results are not clear-cut. Lynn (2001) and Singer (2002, p. 171-172) tested the assumption that an incentive is mainly effective because it increases the motivation of the interviewer. Their results indicate, however, that the effect is most likely directly on the respondent. It would be interesting to repeat this experiment for advance letters and find out whether advance letters have a direct positive effect on survey cooperation or an indirect effect because they enhance interviewer morale and confidence.

Additional support for the assumption that experienced interviewers use tailoring all the time comes from their frustration from three types of refusal. Firstly, interviewers hate having the door slammed in their face before they can start their persuasion strategy. Secondly, they resent more highly educated respondents just saying 'no' and not feeling any obligation to explain why they will not cooperate. And finally, they feel brushed off when the selected individual makes his secretary call the fieldwork organization to inform them that he will not cooperate. In each case, the respondent refuses to enter the role-playing in which the interviewer cajoles the respondent into cooperating by opening up the box of tricks and finding out what appeals most to this person.

Tailoring and social exchange strategies may not be entirely appreciated by non-response adjusters. They would love to have a fixed response propensity for each member of the population, or if necessary a response propensity dependent on survey topic, survey mode, interviewer and response burden, or at least information on the strategies that have been used on the doorstep. This could help them to correct for bias and compute nonresponse weights. Resourceful interviewers, that flexibly cater their approach to the need of individual sample persons, might achieve high response rates but may render statistical modelling more difficult.

The question remains as to whether there is a hard-core group of nonrespondents for whom no amount of tailoring seems to have any effect. Interviewers seem to think such persons exist, and even fieldwork organizations consider some sample persons impossible. Schnell (1997, p. 190), however, concludes from his study of the empirical literature that there is no empirical evidence for the existence of a hard core of adamant refusers who never cooperate in surveys. He attributes the belief in a hard core of refusals to a fundamental attribution error, as observers tend to ignore situational factors that might explain why in this case persons do not respond. Again Goyder, (1987, p. 187) on bias: Part of the explanation for response behaviour rests on people's assessment about the merits and demerits of surveys in general, but as

much, perhaps more, is unique to the particulars of each survey. It is this complexity that makes global remedies for nonresponse such as a universal weighting system to correct bias error-prone, even though there are various useful techniques for use on a survey-by survey basis.

4.4 *Are nonrespondents bad?*

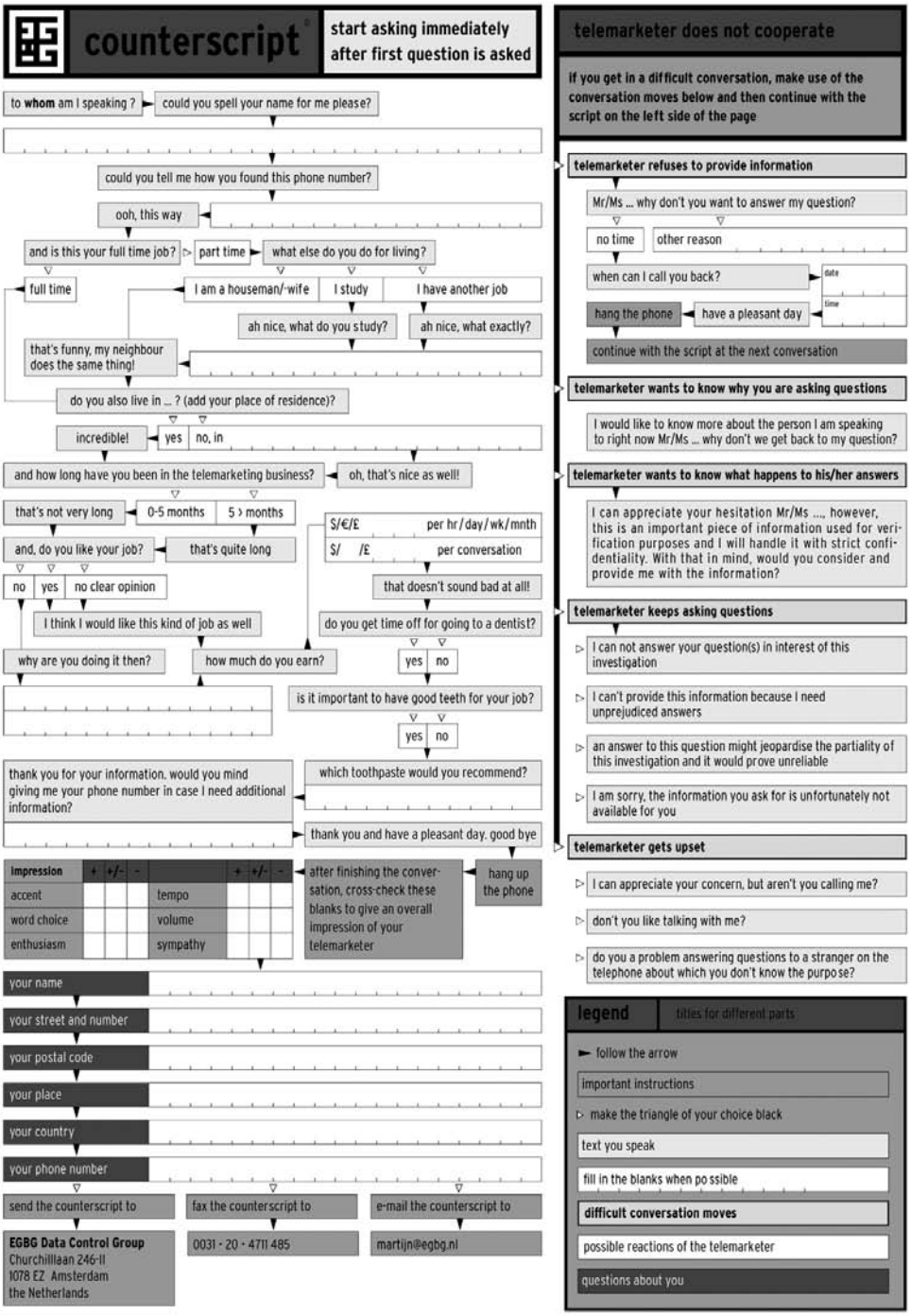
In the present and preceding chapter nonresponse has been discussed from different points of view. Section 3.2 highlighted the problem of establishing contact with sample persons, and described those groups in society that are harder to contact and thus require increased field efforts. Section 3.3 identified those groups in society that are generally considered to be less likely to cooperate. Section 4.2 introduced social isolation and social involvement as possible determinants of survey cooperation, and showed how survey attitudes are related to response behaviour and that survey-related nonresponse most probably has more detrimental consequences than 'accidental' nonresponse. Section 4.3 presented proven techniques of enhancing response rates and ways of tailoring the approach to different groups of difficult respondents. So far, this sounds fairly simple. It becomes more complicated, however, when one looks back and sees that elderly people, who are often at home, more often refuse to participate and that the reason for this refusal can be explained by their more often being single adult households, their fear of crime, their distrust of strangers and their not being accustomed to form-filling, and can be counteracted by their sense of civic duty and their appreciation of receiving attention and a small incentive. It is also made more complex when one sees that younger people who are less often at home, might be interested in a survey which for them is a new experience but might be impervious to dull government letters trying to impress them with the need to provide information on boring policy issues.

Meta-studies aim at identifying individual factors in survey participation and that is why the studies of Goyder (1987), Hox and De Leeuw (1994) and Groves and Couper (1998) are so important. That is also why one should not conduct surveys as a routine activity, but should instead prepare them well, consider possible bias due to the topic and nonresponse from specific groups and try to tailor the survey design. From the overview of survey costs and benefits, social exchange theory and strategies to enhance survey cooperation, it can be deduced that it is worthwhile to study different reasons for refusal, by recording reasons for refusal, recording general doorstep reactions from respondents and nonrespondents and to conduct experimental studies into attitudes towards surveys as some reasons may cause more bias than other reasons. Furthermore, it will be clear by now that there is no single way to the respondent's heart, as different people experience different costs and benefits from a survey and cooperate for different reasons. This implies that tailoring, not only on the doorstep but in all stages of obtaining cooperation, could be a useful approach, and that in nonresponse analysis different reasons for refusal should be distinguished.

In section 4.3 reasons why people do or do not cooperate and strategies to enhance response rates were framed in the theory of social exchange. However, social exchange theory is more a model than a theory that can be refuted (see Goyder, 1987, p. 166).[3]

As social exchange theory is about perceived costs and benefits, it may not lead to testable hypotheses, as many survey aspects can be seen as both cost and benefit. Furthermore – and now we are getting closer to the question of whether non-respondents are bad – the discussion of costs and benefits in this chapter hides from the view a dividing line between two conflicting nonresponse schools. Social exchange theory has been made popular by the work of Dillman and Goyder. In the preceding section many response-enhancing strategies were squeezed into their social exchange framework that might not necessarily meet their approval as they (Dillman, 1999) emphasize the importance of respect for respondents and (Goyder, 1985, p. 25) make a sharp distinction between heartless behaviourists, who seem bent on cajoling and coaxing all these stupid 'sample cases' who don't know what is good for them, into responding, and the believers in social exchange who feel that every citizen is perfectly able to make up his or her own mind on whether or not to cooperate. He gleefully mentions a case of the respondent taking over – much to the anger of the interviewer – as described by the columnist Art Buchwald in 1984, and would probably be highly amused by a recent illustration of the unfavourable attitude towards survey research in the Netherlands (www.egbg.nl) in which the Dutch artist Martijn Engelbrecht presents a 'counterscript' comprising questions to be asked to telemarketers – and presumably to telephone surveys (see couterscript on the next page).Groves and Couper seem to belong to a different nonresponse school, although they too (Groves, Cialdini and Couper, 1992, p. 481) acknowledge the position of respondents: 'One obvious value of this sort of exchange analysis as applied to survey methods is that it acknowledges that nonrespondents can have well-founded rationales for not cooperating with a survey request. In contrast to the view that nonrespondent actions are in some sense based on ignorance, the exchange approach attempts to identify costs and benefits of responding from their perspective'. In their conceptual model on survey participation, Groves and Couper (1998, p. 131) do discuss social exchange theory, but it is hardly central.

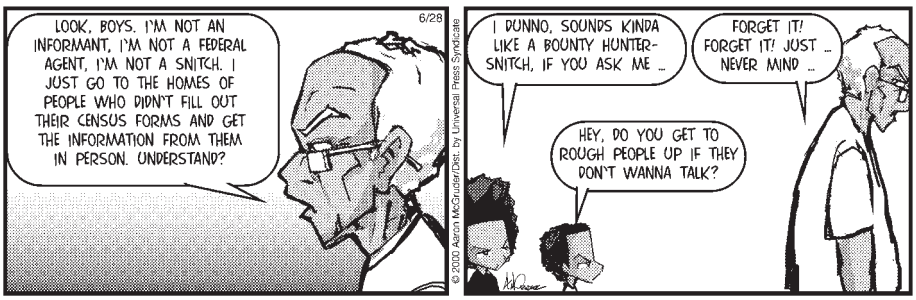
3 Goyder (1987, p. 164, 177) cites Gallegos' (1974) master's thesis in which she states on page 12 that '... Possibly follow-up is so highly effective because in some way it rewards the respondent as well as the researcher. The respondent receives the personal attention of the researcher, who considers him important enough to contact again an again. The implication is that his response is valued and needed; he is essential to a scientific study, and on page 13 that the ... continuing follow-up by the researcher might also be interpreted as imposing a cost on the recipient ... With every additional contact there is a growing implication that the respondents has failed to behave normatively – he has not played fair in the context of exchange.' Groves and Couper (1998, p. 126) infer two opposite hypotheses from social exchange theory about differences in survey cooperation by different SES groups (see section 3.3.3).



They appear to set greater store by the concept of social isolation (section 4.2) which they feel is related to social exchange, as people who are socially involved in their society have more to give to and more to receive from this society. People who are not in touch with the mainstream culture of a society, might be more prone to reject survey requests. This idea should infuriate Goyder (1987, p. 16): 'A sampling suffices to convey the flavour of the literature, however. In all, the profile of the uncooperative respondents identifies precisely the group who, in many sociologists' eyes, forms the most maladjusted and missocialized segment of a society. ... Implicitly, then, nonrespondents have been conceived as deviants too selfish or ignorant to perform their duty when approached by researchers.' At the end of 'The Silent Minority' he comes back to this point, though somewhat more benignly (p. 187): 'My considered view is that, cumulatively, nonresponse has meant that the socially central and active, perhaps behaviourally manipulable and exchange-oriented, citizenry have tended to be overrepresented on surveys ... The very multi-dimensionality of survey nonresponse helps counteract the bias, however, so that one component of nonresponse at times helps nullify another, as seen in the profile for noncontacts versus refusals ...'

Leaving aside opposition between behaviourists with a deterministic view on selling surveys and social exchange theorists who believe in the free volition of citizens, their ways of studying nonresponse processes and underlying factors and their suggestions about increasing response rates are remarkably similar, so we will leave them side by side in section 4.3. This might be a good place to mention again the paradox described by Vuijsje in section 1.3, namely that citizens expect the state to take care of their well-being through subsidies and services, but feel they have every right, are even commendable, to see requests for information from the same state as unacceptable infringements of privacy. Now that we have decided that nonrespondents are not 'bad' in two senses - they have every right not to cooperate in a survey and are not necessarily part of a maladjusted part of society - the next question is whether survey researchers are bad. They cajole (or pester) sample persons into cooperating, re-approach sample persons even after they have refused and want to know the ins and outs of respondents' private lives with - in the best case - only a trinket in return.

The paradox is that response-enhancement strategies may have the simple function of showing respondents that their opinion really counts. Re-approaching refusers can appeal to the scarcity argument mentioned earlier and indicate that it is really important to the sponsor, the fieldwork organization and the interviewer that this particular person cooperates. This might be one of the reasons why refusal conversion seems to work. This leads to the uneasy compromise that respecting the free will of citizens whether or not to cooperate, is a good thing, but that taking measures to enhance response rates is even better. In doing this, one should bear in mind that



undue pressure will in the long run poison the survey climate, and that the ethics of survey research should be kept in mind (see Goyder, 1987, p. 34 and Singer et al. 1993 and 2003). And in trying to obtain cooperation, respect for the respondent is key. In the end tailoring should not be seen as a bag of tricks but as a responsible way of adapting the fieldwork approach to the preferences and interests of the respondent.

5 Studying nonrespondents

5.1 Introduction

In the earlier chapters several approaches to studying nonresponse have been presented.[1] Section 2.2 mentioned the importance of sampling frame data. Previous information from sampling units can either come directly from the sampling frame or from low-level geographical databases that can be linked to the sampling frame and contain information on the size and composition of the population, dwelling types, neighbourhood SES, etc. Paradata or information on the fieldwork process as indispensable tools in analysing nonresponse were discussed in section 2.6.4. Paradata also comprise reasons for refusal (section 4.2) and interviewer observations of sample unit dwellings and neighbourhoods which provide similar information to low-level geographical databases, but are generally more specific.

Other approaches to studying nonresponse include studying panel nonrespondents, recording doorstep interaction, asking survey participants questions about the current survey , mounting surveys on surveys and conducting experimental studies. The results of these approaches were presented in section 3.4. This chapter introduces three additional approaches, namely comparing reluctant or late, or in general 'difficult' respondents with easy respondents, assuming that the former are more like final nonrespondents than the latter, and asking questions to nonrespondents that are willing after all to provide (some) information. This can either be done by requesting each nonparticipant to at least answering one central question, or by re-approaching a sub-sample of refusers in a follow-up survey among persistent refusers.

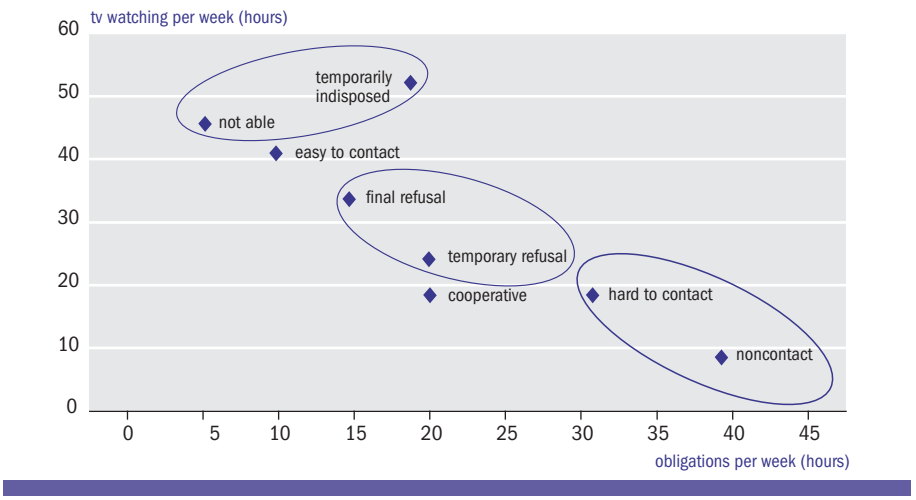
Section 5.2 discusses theories based on the assumption that difficult respondents are more similar to final nonrespondents than easy respondents. Section 5.3 gives an inventory of the results of extended field efforts, aimed at bringing in at least a part of these difficult respondents. Section 5.4 follows another course in nonresponse research, i.e. collecting substantive variables from all refusals, or a random sample, rather than enhancing response rates per se. Section 5.5 will look back on the literature presented in the first four chapters of this study and outline how these results will be used in the subsequent empirical chapters.

1 The overview of approaches to studying nonresponse is based on Groves and Couper (1998, pp. 49-50).

5.2 *Continuum of resistance and classes of nonparticipants*

Nonresponse researchers appear to agree with Schnell and Goyder (section 4.3.5) that there is no hard-core group of nonrespondents, or at least not a large group. This can be inferred from the general assumption that by increasing funds and field efforts or manipulating the design, response rates can be increased substantially. Furthermore, they generally assume that participants at whom the greatest efforts were directed are similar to nonparticipants who would have required even larger efforts. This is the rationale behind both studies on the effect of increased field efforts and studies of the differences between classes of respondents. The first approach is qualitative and assumes that there are different types of nonparticipants that are similar to different types of participants. The second approach is quantitative and assumes that more extensive field efforts would have resulted in higher response rates, and that the most difficult respondents would have been nonparticipants had the efforts been less. The landmark paper that distinguishes these two approaches is Lin and Schaeffer (1995) who describe two simple, ad hoc methods for assessing the impact of nonparticipation on survey estimates, the *classes of nonparticipants* model and the *continuum of resistance* model.

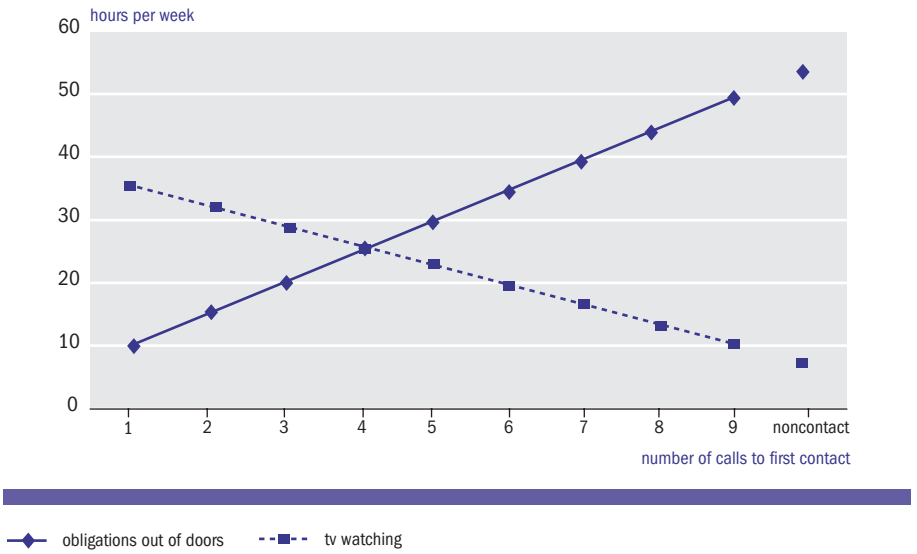
Figure 5.1a Hypothetical relationship between response and nonresponse classes, obligations out of doors (work, study) and watching TV: ideal model of classes of nonparticipants



The first qualitative model distinguishes between *classes of nonparticipants* such as noncontacts and refusals. Specific classes of participants, e.g. temporary refusals and hard-to-contacts, can be used to estimate the impact of different types of nonparticipation, e.g. final refusals and noncontacts. Other potential classifications distinguish between nonresponse due to situational factors and for more permanent reasons (propitiousness and inclination (Smith, 1984)) or between active and passive

nonrespondents (Rogelberg et al., 2003), both discussed in section 4.2.2. Most studies, however, distinguish between three classes of participants, namely temporary refusals, hard-to-contacts and easy respondents. A graphical representation of the model is given in figure 5.1a. It presents the hypothetical position of classes of participants and nonparticipants with regard to hours per week spent on work and study out of the house (x-axis) and hours per week spent in front of the TV (y-axis). In this example those who were temporarily indisposed but participated are similar to those who were not able to respond because of illness, temporary refusals are similar to final refusals and hard-to-contact respondents are similar to noncontacted respondents.

Figure 5.1b Hypothetical relationship between contactability, obligations out of doors (work, study) and watching TV: ideal model of continuum of resistance



The more quantitative continuum of resistance model introduced by Lin and Schaeffer '... assumes that nonparticipants are all similar to each other and that if nonparticipants took part in the study, they would respond like a particular group of participants – usually those participants who were difficult to contact for the interview'. The graphical representation of the model in figure 5.1b illustrates the hypothetical relationship between number of calls to first contact and the hours per week spent on obligations outside the home (work, study) and watching TV and shows that by making more calls and thus contacting more sample persons information becomes available from respondents that are quite similar to nonrespondents in terms of hours spent in front of the TV and hours spent on work or study. The nonresponse models that have been used in the literature are usually similar, though generally not identical, to the continuum of resistance model. In the end it will turn out that often a combination of both models is employed. This section will give a short theoretical overview of related models and focus on the problems in

measuring resistance. The next section will present empirical results of studies in which extended field efforts were deployed.

Resistance is sometimes related to contactability as Lin and Schaeffer suggested, sometimes to reluctance to cooperate and sometimes to an unspecified difficulty. Rather than breaking the resistance of the respondents by increasing the number of calls, one could break their resistance by increasing the size of the incentive, or through multiple refusal conversion attempts by increasingly experienced and success-ful interviewers. Alternatively, the resistance to cooperate might be measured by the attitude of the respondent towards surveys, the time needed to persuade the respondent to participate, the number of negative comments before cooperating and the interviewer assessment of the likelihood that a refuser will ever cooperate. Earlier, Smith (1983) introduced the 'difficulty method', which seems fairly similar to the continuum of resistance.[2] He gives three reasons for difficulty: availability (being at home at a given time), contactability (which he defines as the probability of some other responsible household member being at home at a given time) and reluctance. Goyder (1987, pp. 14-15) elaborates a continuum of resistance to cooperate in his '... wave design in which 'early' respondents – those who respond with little prodding – are compared with later ones.... The problem now becomes, however, whether 'reluctant' respondents (who may simply be inaccessible) furnish reliable information about the terminal nonrespondents who might include, in unknown proportion, determined resistors'.

Reluctance seems much more difficult to measure in terms of resistance than accessibility. Smith (1983, p. 391, note 2) '... If temporary refusers were further subdivided into easy or hard to convert, or some other refinement, such as attempts needed to convert, then it would be possible for other relationships to emerge. Apparently, no one has attempted such a refinement and the usually small number of total converts would make such a refined analysis difficult.' Later he observed (Smith, 1984, pp. 398-399) 'Although we can use respondent characteristics to predict difficulty in completing an interview, these characteristics cannot be used to estimate the attributes of nonrespondents. This is because the number of attempts measures primarily how accessible a person is, while the final nonrespondents are made up primarily of refusals, not inaccessibles.'

The continuum of resistance models are built on three assumptions. Firstly, response propensity should be correlated with survey outcomes. If easy to contact respondents do not differ from hard-to-reach respondents, aiming for a high contact

2 'Difficulty may be reflected by the number of mailings, visits, or telephone calls, how long it took to get a response, or some measure derived from these indicators of difficulty.... If a linear or some other regular relationship is found, then this association is used to impute the distribution of the variable among nonrespondents' (Smith 1983)

rate would have no effect on bias, nor would aiming for a high cooperation rate if temporary refusals do not differ from the initially cooperative. If this were the case, the only rationale for improving response rates would be that more respondents would increase precision, and that rather than drawing a larger sample to obtain more respondents, one might as well make more calls to a smaller sample (resulting in a higher response rate and more respondents). The second assumption is that final nonrespondents are similar to hard-to-reach respondents, who will either be persons requiring many calls or converted refusals. The first aspect is less pertinent, as low noncontact rates of 3% can be achieved in good surveys. The second aspect will be hard to test, as the final refusal rate is generally high (25-30% in well-run studies), whereas the percentage of converted refusals will often be much smaller. The final assumption is that resistance can be measured. This is problematic for three reasons. Firstly, this requires interviewer efforts to be meticulously recorded. Sadly, this is not always the case. In addition, resistance will not necessarily be a characteristic of sample persons, but also be related to the field approach and doorstep interaction. Interviewers who call at convenient working hours may end up with a large number of hard-to-contact respondents who would have been very easy had they been approached during evening hours. Lynn and Clarke (2001, p. 8) compare two surveys, one of which had more difficult to contact households than the other. They ascribe this to less efficient calling strategies in the survey with more difficult households, due to heavy workloads on interviewers and field staff at that time.

An additional problem with measuring resistance in terms of reluctance is that the assessment of the interviewer might be a self-fulfilling prophecy. As emerged from the German ESS fieldwork (translation Stoop): 'We should mention here, however, that a categorisation from sample persons in refusal groups (such as 'soft' and 'hard') by the face-to-face interviewer in the field, may be diverse and be open to differences in interpretation' (Infas, 2003, p. 25; see also Schnell, 1997. p. 151). Interviewers may not really expect 'their' refusers to cooperate, which might explain the success of deploying a new interviewer who is not hampered by previous failures to obtain cooperation. Kennickel (2000) seems to be rather concerned about the idea of having interviewers assess the likelihood of the success of re-approaching a particular sample unit[3];

3 'The deployment of interviewers has a strong effect on the timing of work, and it almost certainly has a strong effect on the rate at which interviewers attempt to complete cases. Unfortunately, the decisions made in allocating field personnel are rarely captured in a systematic way that would be useful for analysis. Moreover, it would be very hard to believe that management decisions affecting respondents' and interviewers' incentives do not have a very powerful effect on the arrangement of work at the level that is clearly visible only to the interviewers. To keep their jobs or to be eligible for bonus pay, interviewers typically have to complete a specified quota of cases. Such standards alone have the effect of steering interviewers away from efforts on cases that may be or may be expected to be particularly difficult' Kennickel (2000, p. 40)

Lievesley (1983, p. 298) had more confidence in their judgment: 'How do we distinguish those people worth revisiting from those who are extremely unlikely to take part? Perhaps the most obvious way is to instruct the interviewers to record the reason for refusal and to make a subjective judgement regarding reissuing based upon this reason. ... selective reissuing on the basis of interviewers' assessments is preferable to random reissuing and we plan to continue experimenting with this procedure.' Thus reluctant respondents who offer the prospect of future success will be re-approached and converted, whereas fewer efforts are expended on reluctant refusers who do not seem to be very promising – condemning them, possibly undeservedly, to remain final refusers.

Brehm (1993, p. 128-130) sees another inherent problem in applying the continuum of resistance to reluctance. He combined a continuum of resistance with respect to accessibility and to cooperation in his approach to modelling the survey process, in which ever more administrative measures (more calls, sending a letter to try and persuade reluctant sample persons, trying to convert a refusal) would increase survey participation. The difficulty he found with this model is that persuasion letters are only sent to reluctant respondents, and therefore seem to have a negative effect (as reluctant respondents more often turn into final refusers and no persuasion letters are sent to respondents who cooperate instantaneously). As he remarks in a footnote (p. 130): 'If one's interest lies in how effective these techniques are .... the persuasion letters and refusal conversions would have to be randomly assigned treatments, not treatments assigned on the basis of an initial refusal.'

Curtin, Presser and Singer (2000, p. 426) mention a pitfall of comparing easy-to-contact and hard-to-contact respondents in a study on nonresponse in the Survey of Consumer Attitudes (SCA): 'Truncating the SCA samples, thereby creating a lower response rate, is not the same as conducting a survey designed to yield a lower response rate. If the SCA were carried out using a limited number of calls, the calls would probably be scheduled differently..... Nonetheless, we believe our simulations provide a useful approximation of response rate differences from different designs.'

Keeter et al. (2002, p. 126-127) present two viewpoints that are not identical to the continuum of resistance or classes model, but which clearly outline the different impact of enhancing response rates dependent on the underlying nonresponse process: 'One viewpoint assumes that efforts to increase response rates will gradually reduce nonresponse bias by bringing into the respondent pool more diverse groups of respondents, obtaining, as a function of the response rate, a progressively more balanced representation of the entire survey population. This implies that distributions of survey variables change as the response rate increases. Another viewpoint assumes that there is a set of characteristics that affect the survey participation decision (e.g., topic, burden, sponsorship, interviewer behavior), and that persons to whom those attributes are not attractive will be nonrespondents. Thus as the

response rate increase, the respondent pool is being increased by more of those persons located and contacted for whom the survey conditions are minimally acceptable. The rest remain nonrespondents. This implies that survey statistics stay relatively fixed as the response rate increases.'

Within the continuum of resistance model, increasing response rates will decrease nonresponse bias. This does not hold for the classes model. In practice, however, the two models are hard to distinguish. The continuum of resistance, as defined by Lin and Schaeffer, assumes that there is a single mechanism behind nonresponse, and that nonparticipants – had they participated – would most likely be similar to those participants who were difficult to contact or reluctant to cooperate. By increasing interview efforts, more potential nonparticipants become participants. Nonparticipation can be caused, however, by noncontact and noncooperation, and the difference between both types of nonresponse was highlighted in section 2.6. Consequently, the increased field efforts should be described in terms of increasing the number of calls and converting refusals. This two-dimensional continuum of resistance generally boils down to the four classes of participants presented in table 5.1, easy respondents, hard-to-reach respondents, temporary refusals (who are easy to contact) and hard-to-reach temporary refusals. The latter two classes are generally merged, if only because the number of converted refusals is often small (section 2.6.3). For the real qualitative 'classes' model the studies by Rogelberg and Smith mentioned earlier are relevant.

Table 5.1 lasses of participants and increased interviewer efforts

	small number of calls to first contact	large number of calls to first contact
one contact	easy respondents	hard-to-reach respondents
more contacts (converted refusals)	temporary refusals	hard-to-reach, temporary refusals

From the studies presented in this section, it appears that the concept of resistance is hard to operationalize and that it is difficult to distinguish between the continuum of resistance and the classes of nonparticipants model, because the terminology and elaboration of studies in this area are far from consistent, and the differences between the models are small if the classes model is based on a dichotomization of an accessibility continuum and an amenability continuum, which is generally the case. Lin and Schaeffer (1995, p. 252) ended their test of both models with empirical support for neither model: '... our analysis clarifies that these methods really are ad hoc: they require a large number of small decisions – such as what to use as an indicator of the level of effort to measure 'resistance,' how to distinguish easy–to-locate from hard- to-locate respondents, or what classes to use – that must be made without

a great deal of guidance from either theory or data but that may greatly affect the resulting estimates of nonparticipation bias.'

The next section will present empirical results of studies of extended field efforts in which difficult respondents are compared with easy respondents without sticking to one of the models presented here. The underlying assumption will always be that extended field efforts will lead to higher response rates, which is good because the extra respondents reduce bias to nonresponse. The results of the studies in 5.3 will show whether these assumptions are valid.

5.3 *The results of extended field efforts*

Empirical studies on the continuum model are not always very convincing, either because there is no information on the most resistant group - the persistent refusers - or because there is inconclusive evidence to enable a decision to be made on how 'resistant' a respondent was, or because there is not really a continuum on which persons are located. Bates and Creighton (2000), for instance, adopt the continuum of resistance model but besides having measurement problems they cannot distinguish properly between late and difficult cases. They end up with non-conclusive evidence, find that targeting the last 5% produces small changes in survey outcomes and conclude that: 'Late outcomes may or may not be reflective of nonrespondents, but their input is still critical to producing unbiased rates and minimizing non-sampling errors.' In other studies, the 'resistance' put up by the respondents seems fairly weak. The much cited study of Guadagnoli and Cunningham (1985) amounts to nothing more than the finding the 128 physicians who sent their mail questionnaire back after the initial request, did not differ substantially from 83 other 'resistant' physicians who did this only after a single reminder. The chronological overview of studies on resistance below may give more persuasive evidence.

Stinchcombe, Jones and Sheatsley (1981) distinguish between classes of participants and found substantial differences between temporary refusals and the easy respondents who were both accessible and cooperative. The hard-to-reach group did not differ from this easy group, so they recommended converting initial refusers rather than revisiting the absent. Smith (1983) found the obvious relationship between accessibility and socio-structural characteristics (those who were employed were less often at home, and in families with small children more often someone was at home), and that the hard-to-reach appeared to be more liberal and more wealthy, and had a more active and youthful life style (having received a traffic ticket, member of youth groups, watch less television, favour legalization of marihuana). Only the lifestyle variables added explanatory power to the basic demographic model. Smith (p. 398) remarks that '... While this means that very few nondemographics are independently related to difficulty, many variables (27% ...) are closely enough related to the independent variables to vary notably with number of attempt. This means

that not only will variables directly related to availability be affected, but many attitudinal and behavioral variables will also be affected.' Bivariate relationships did not seem to vary with difficulty. Furthermore (pp. 399-340) he found minor differences between cooperative respondents and temporary refusals and identified 'urbanness' as the only variable having a strong relationship to temporary refusals. He suggested the following explanations. 'If we accept temporary refusals as indicators of final refusals, then the lack of significant associations' (between hypothesized variables and cooperative/ reluctant refusals) 'suggests that except for city type, final refusals are not significantly different from completed cases and therefore little bias is introduced.' This might mean that the variables expected to be related to survey behaviour were not the right ones, or that refusal is a random occurrence. He also observed that it might be true that temporary refusals do not adequately indicate attributes of final refusers.

Clarridge and Massagli (1995) analysed two surveys among physicians. In the first the response rate was enhanced from 50% to 82% by additional field efforts (the final part of which was directed at a subsample of refusers, of whom 50% cooperated). The second had a response rate of 83% and paradata were collected on the efforts expended (letters mailed, faxes sent), time in the field, and number of telephone calls made in pursuit of the interviews. The first study showed no differences in background variables and some differences in survey outcomes. In the second study, reluctant and hard-to-contact respondents were (sadly) pooled. The authors did not find compelling evidence across variables that difficult respondents were systematically different.

Smeets (1995) presented three models for weighting for nonresponse, namely the Politz-Simmons adjustment[4], the difficulty approach and the converted user approach. She then rejected the first and the last and applied a variant of the difficulty approach as introduced by Smith (1983) (see above). She distinguished three groups, namely those who agreed immediately, those who were initially absent or were detained by other causes, and those who initially refused to participate and had to be converted. This is actually more like the classes model than Smith's difficulty method, which assumes a linear or at least a one-dimensional relationship. Smeets found differences between groups, notably that converted refusals were politically less interested than the other respondent groups, but the effects were rather small and her results were far from consistent. She concluded that converted refusers might not be similar to the final refusers after all. Van Leest and Burhenne (1997) analysed differences

4 In the Politz-Simmons approach only one single randomly timed call is made to each sample unit. Completed interviews are then weighted by the probability of finding the sample person at home. This probability is estimated by asking a number of questions on the respondent's being at home at different times during the last week. Groves (1989, pp. 169-172) is fairly critical about this approach, because it only addresses nonresponse because of noncontact, asks about being at home in the previous week rather than during the fieldwork period, requires the respondent to be accurate in the reporting of at-home patterns and requires randomly timed calls.

between easy-to-reach respondents, respondents requiring many calls and nonrespondents in the AVO 1995 survey. They used neighbourhood data from general postcode databases and from interviewer observations of neighbourhood and residential characteristics, and population data from Statistics Netherlands and the GfK MiniCensus (GfK, 1999). City dwellers, people living in apartments, people aged 18-34, single people and people with no children, the highly educated and people in work required many calls (to reach and/or to convert them). The higher than usual number of calls in this survey (4-14, whereas three calls were made in earlier rounds) substantially increased the presence of hard-to-reach groups in the sample, with the exception of inhabitants of large cities. City-dwellers, people living in apartments and elderly households refused to cooperate relatively often. Comparisons between early respondents (1-3 calls in 1995), late respondents (4-15 calls in 1995) and the less rigorous 1991 survey (3 calls maximum) showed major differences in net sample composition. The increased number of calls made the net sample more similar to the population. With respect to substantive variables (use of services), however, the majority of the differences disappeared after weighting for socio-demographic variables, except for going to the movies.

Voogt, Saris and Niemöller (1998, p. 254) introduced a combination of the continuum and classes model, and placed the classes (or rather individual members of the population) on a response-continuum that ranged from 'will never respond' to 'will always respond'. Both temporary refusals and difficult-to-reach respondent would be closer to the 'will never respond' side. Surveys can also put on this continuum, based on the quality of the interviewers, the maximum number of calls and the maximum number of refusal attempts. In their analyses they focus on the classes aspect and ignore the continuum. They compared those who were interviewed without problems, the temporary refusals and the difficult to reach, and found significant differences with respect to age, education and sex, but not with respect to urbanization and political interest. In their study the correlation between political interest and voter turnout was highest among the temporary refusals.

Aarts and Van der Kolk (1999) and Van der Kolk (2001) analysed differences in voting behaviour in the Dutch Parliamentary Election Study (DPES) 1998 between respondents in the face-to-face survey, nonrespondents who were willing to answer a very short telephone questionnaire and final nonrespondents.[5] The response rate was approximately 50%, and 44% of the nonrespondents gave some information on the phone. Response rates were lower and refusal conversion was less successful in urban areas.

5 This approach can also be seen within the framework of the Central Question Procedure (see the next section).

The authors tried to find out whether the converted refusals were similar to the final refusals, which was complicated by the fact that the survey comprised different waves and that the pre-election interview might affect voting behaviour. Their results are not entirely conclusive. They found that turnout was higher among converted refusals than among both non-contacts and final refusals, but suggest that this might partly be due to the pre-election interview effect. The conclude (p. 166, translation Stoop) 'This implies that we may after all be able to use the data we collected on nonrespondents in a telephone interview to say something about the entire group of nonrespondents.'

Curtin, Presser and Singer (2000) analysed the relationship of the Index of Consumer Sentiment (ICS) to number of calls and to initial cooperation versus refusal, controlling for time of the survey. Respondents who did not require refusal conversion had higher ICS scores and were more optimistic about the economy, possibly because refusal conversions mainly came from lower socio-economic strata. Those interviewed on earlier calls had lower ICS scores, possibly because the hard-to-reach were disproportionately of higher status. The hard-to-reach were younger and more affluent, and the reluctant respondents older and less well educated. After controlling for these demographics, the association between reluctance and ICS disappeared, and that between accessibility and ICS was reduced. With respect to estimates of change (the main purpose of the ICS) the effect of amenability was nonexistent, and that of accessibility small. They also analysed what would have happened if interview efforts had been smaller, by comparing five groups: initial cooperators, respondents interviewed after 1-5 calls and after 1-2 calls, and initial cooperators interviewed on 1-5 calls and on 1-2 calls. Due to an ingenious design these groups were independent. They found effects on the level of ICS only for the 1-2 calls groups, and virtually no effects on the change in the ICS. Summarizing, differences in ICS seemed to be mainly due to the composition of response groups rather than to bias, and were mainly related to contactability rather than reluctance. Effects on the level of ICS could be observed, rather than change in the ICS. The added value of extended interview efforts seemed hardly worthwhile. They mention, however, that these results refer to relative rather than absolute nonresponse bias, as they discuss the effect of an increase of response rates from 60% to 70%, whereas the size of the nonresponse bias due to the final refusers in unknown.

Borg (2000) studied the difference between early and late respondents in an e-mail survey among employees of a large organization. One problem he draws attention to is that early and late respondents in mail surveys are often hard to distinguish from each other. Respondents may be late because they simply forgot to return the completed questionnaire, or the mail could have been late. His e-mail survey, preceded by a very intensive information and motivation campaign, would be a very good way to distinguish between early and late respondents, and even to extrapolate the results of the latter to final nonrespondents, as the start time of the survey was clear and the incoming flow of responses could be measured precisely (with one interruption caused by the 'I love you'-virus). In Borg's study accessibility plays virtually no

role. His results show a response rate of 80% and no difference at all between early and late respondents, either in a theory-based analysis or in a fishing expedition (not even with regard to commitment to the organization or satisfaction), with one exception: employees in American and non-German European countries, in Japan and China were rather late, a number of Asian countries had obviously made a point of achieving a response rate of 100% during the first day, and Germans were equally spread across early and late respondents.

Lynn and Clarke (2001) compare the effect of extended field efforts in British surveys after separating for non-contact bias and refusal bias. Respondents in hard-to-contact households were younger and more likely to be employed. Furthermore these respondent were more often heavy smokers and drinkers and more healthy, which might be related to age and employment. They conclude that hard-to-contact respondents differ from the others, whereas reluctant respondents are actually quite similar to the initially cooperative. Lynn et al. (2002, p. 142) conclude from similar analyses that '.. it is the difficult to contact that are most different from the easy to get. This would appear to suggest that resources for extended efforts might be better concentrated upon making contact with difficult to contact households than upon attempting refusal conversion.' They make one reservation, however, because the final contact rate was very high in the British studies (96.3% to 98.2%) whereas about a third of the sample remained as refusers. It might be that 'although converted refusals are similar to the easy-to-get, the remaining ('harder') refusals are rather different'.

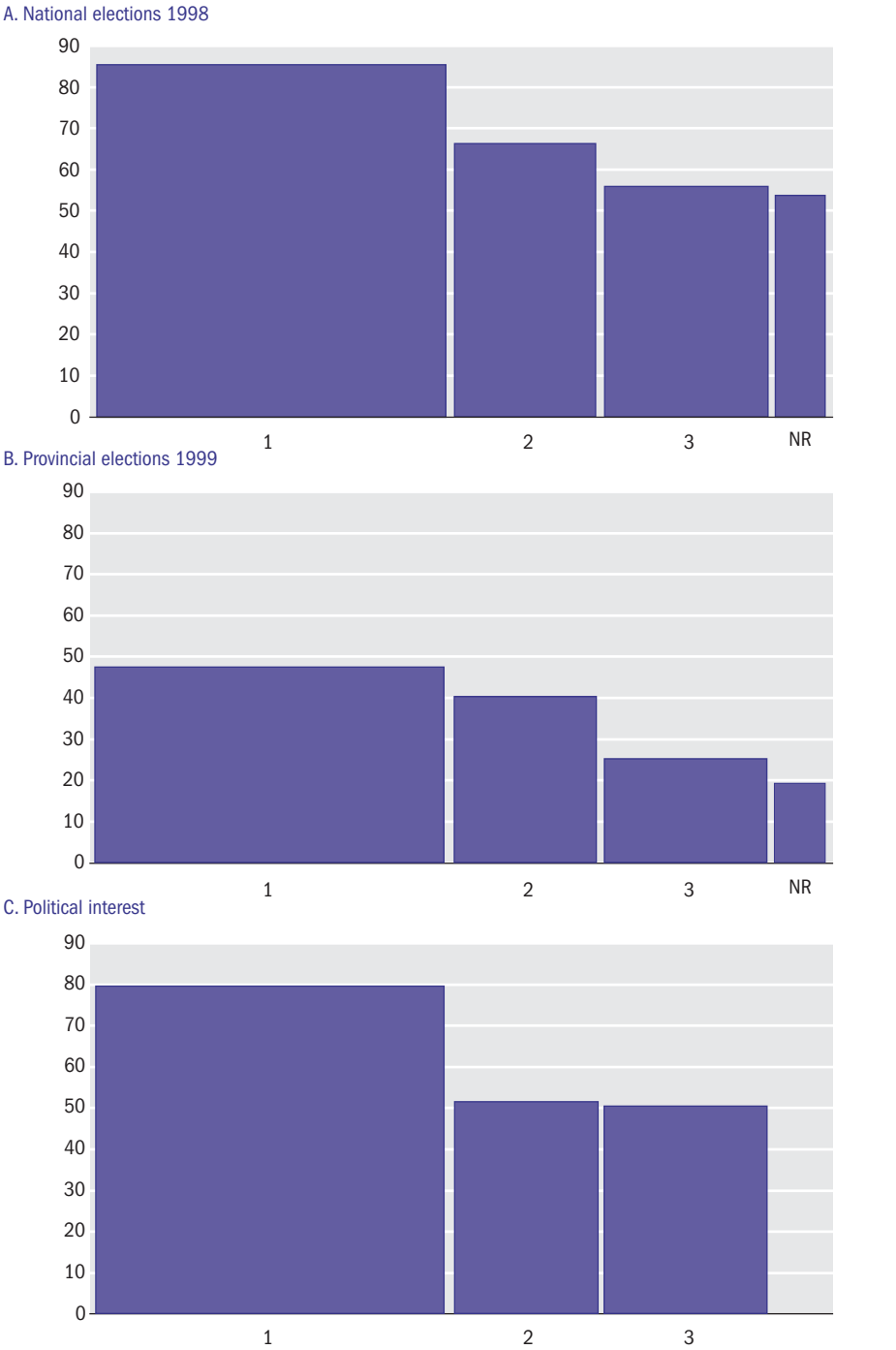
Keeter et al. (2002, p. 142), in their analysis of the consequences of reducing non-response in a telephone survey, compared hard-to-reach with easy-to-reach households, and reluctant with amenable households. They found that the number of calls to first contact with a household is significantly correlated with age and education, the young and the better educated being more difficult to reach, as were households containing one adult and those with employed respondents. Many other significant differences on media use, political knowledge, social integration, were strongly correlated with age and education. In the analysis of reluctance, they ran into the following problem: 'For all cases except the one-person households we do not know whether the same person refused and later granted a completed interview; in many instances different people will have been involved. Accordingly, we refer to 'amenable households' and 'reluctant households'.' Not surprisingly, amenable households were rated by interviewers as more cooperative, more interested, and more likely to participate in future surveys. A number of survey outcomes differed significantly, but not more than could be expected by chance. Reluctant households appeared to be less racially tolerant, however, but that may partly have been due to interviewer staffing differences.

Teitler, Reichman and Sprachman (2003) studied the costs and benefits of improving response rates for fathers in fragile families. They distinguished between participants

according to interview mode, namely in hospital (soon after the child was born), by telephone when in hospital was not possible, and face-to-face (if the telephone approach also did not succeed) and according to interview costs (which were clearly higher after each successive type of approach). As they mention, interview mode and costs as measures of effort make their study difficult to compare with other studies in which the more convenient metric 'number of attempts' was used (which was not available). Their study is special because they had information on all responding and nonresponding fathers from the mothers. They found that the most difficult cases (either according to mode or to costs) do not appear to be similar to the final nonparticipants. The highest effort cases not only failed to reduce response bias but were also poor proxies for nonrespondents. Teitler and his colleagues clearly found a point of diminishing returns beyond which an increase in response rate did not add to the quality of the data. Resources could have been better used in other parts of the project. They acknowledge, however, that it will not be clear at the outset when this point of diminishing returns has been reached. Rather than just settling for low response rates, they encourage fieldwork organizations to take advantage of modern technology and the availability of real-time data to make informed and practical decisions about resource allocation. They even conclude that in their study, even at very high response rates, significant nonresponse bias remains. They end with the following remarks (p. 137): 'As important it is for projects to allocate data collection resources efficiently, this strategy will help us neither adjust for or prevent nonresponse. To adjust for nonresponse, we need information about how nonrespondents and respondents differ. To prevent nonresponse, we need to learn more about the cognitive processes underlying survey participation in today's world.'

Voogt (2004, pp. 89-94) gives an inventive and persuasive example of a mixed continuum of resistance model in a study in which very high response rates were achieved. He started with a net sample of 1000 Dutch inhabitants of the town of Zaanstad whose electoral behaviour was known. All sample units with a known telephone number were contacted by telephone, the others received a mail questionnaire. Group 1 (51%) were the persons who cooperated at the first try. All refusers in the telephone interview were contacted again to persuade them to cooperate, and mail non-participants received a second mailing with a letter emphasizing the importance of the study. In this second try, reluctant sample persons were given the option to answer only two questions, on voting behaviour in the national elections and on interest in politics. Those who provided the requested information at this stage were group 2 (19.7%). The remaining refusers, those who could not be contacted by telephone and those who did not return the mail questionnaire, received a personal visit from the interviewer. Those who cooperated in the face-to-face situation were group 3 (22.4%), the remainder were nonrespondents (6.8%). Figure 5.2 presents the combined results for all interview modes and shows that the respondents who required most efforts are clearly more similar to the final refusers with respect to voting behaviour. The political interest of the final refusers is not known; here the only difference is between easy respondents and the rest.

Figure 5.2 Continuum of resistance model, voting and political interest, Voogt, 2004



In summary, the conclusion of this overview appears to be that it is useful to distinguish between type of resistance (accessibility or amenability), that hard-to-contacts seem to differ more from easy-to-contacts, and that temporary refusals are not that dissimilar from the initially cooperative (with the exception of the early study of Stinchcombe, Jones and Sheatsley (1981)), that noncontact rates can be minimized, that the added value of increasing response rates when they are around 70% seems questionable, that temporary refusals are not necessarily similar to final refusals and that the possibility cannot be ruled out that final refusal causes substantial nonresponse bias.

The fact that increasing response rates from 60% to 70% or from 70% to 80% is very costly and may not reduce bias, the fact that even at this level substantial nonresponse bias may remain, and the fact that pushing the initially reluctant respondents to cooperate may irritate respondents and have detrimental effects on data quality (Voogt, 2004, p. 81) and may not be legitimate in some countries, leaves plenty to worry about. The next section will discuss possible alternatives.

5.4 *Measuring nonresponse bias*

Converting refusals may not necessarily reduce bias, using converted refusals as proxies for final refusals does not necessarily lead to better estimates. This casts doubt on the benefit of efforts to increase response rates by minimizing final refusals. Other empirical results too make one wonder whether high response rates are better. Keeter et al. (2000) compared the results of a rigorously conducted survey with a standard-light version and found few differences. Merkle and Edelman (2002) concluded from an analysis of nonresponse in exit polls that there is no relationship between nonresponse rate (here mainly due to refusal) and error. They feel their results 'buttress Krosnick's (1999) conclusion that 'the prevailing wisdom that higher response rates are necessary for sample representiveness is being challenged' and that 'it is no longer sensible to presume that lower response rates necessarily signal lower responsiveness' (p. 255). They also cite a case where a slight increase in response rates significantly increased bias and end: 'This should serve as an eye-opener to those who pursue higher response rates for their own sake without considering the impact of survey error. Devoting limited resources to increasing response rates with little or no impact on survey error is not money well spent, especially when that money might be better spent reducing other sources of error' (see Groves, 1989).

As enhancing response rates may not necessarily minimize bias nor provide data for estimating bias, other techniques will be discussed here that aim at studying bias without 'interfering' with response rates, namely a follow-up survey among refusers and the central question procedure. These techniques have been developed with the explicit aim of minimizing nonresponse error and providing data for correcting nonresponse error by adjusting for refusal. Here, the methods will be presented in a

more descriptive manner, as techniques for discovering whether and how nonrespondents differ. Surprisingly, neither technique is used very often, at least not in the Netherlands.

In 1946 Hansen and Hurvitz outlined the strategy of double sampling as an adaptation of the principles of double sampling from Neyman. Their starting point was the high nonresponse rates in and low costs of mail surveys, and the high costs of personal interviews, which at that time appear to have elicited a substantially higher response rate. Double sampling implies approaching the complete sample using an inexpensive survey mode and settling for moderate response rates, and subsequently drawing a sample from the nonrespondents, approaching these with an expensive mode and achieving complete cooperation among this subsample.[6] Hansen and Hurvitz presented an overview of sampling designs that shows that – dependent on the cost difference between the inexpensive and the expensive mode – interviewing a sample of the nonrespondents on the first sample, yielded the same standard error of the mean as following all nonrespondents at substantially lower costs.

Groves (1989, pp. 165-169) discusses Hansen and Hurvitz's model as a useful way of minimizing refusal bias, possibly in combination with the Politz-Simmons approach for correcting for being at home. As drawbacks he mentions that the Hansen and Hurvitz' double sampling procedure focuses on sampling error, and might be less cost efficient if nonresponse error, the difference in survey statistics between respondents and nonrespondents, is small. Furthermore, the approach assumes 100% response rates in the follow-up sample (although Groves still considers it a useful procedure if the nonresponse rate in the follow-up is lower than in the original sample). Finally, the approach ignores measurement errors due to mode differences. Still, he considers the approach to be a useful framework for the elaboration of survey cost and error models. Bradburn (1992) considers the double sampling approach and other methods in Groves (1989) book as the response to the nonresponse problem: 'These methods share one element in common, that is, learning more about the characteristics of hard-to-get-respondents and, with some additional efforts, even the nonrespondents. The data on these difficult respondents can then be used to do better weighting that can incorporate more information into the weighting procedure. While we may not have the resources to get high response rates across the board, we can allocate the data collection resources in a more targeted manner to learn more about the possible bias arising from low response rates.'

Elliot (1991, pp. 38-40) introduces double sampling as a standard procedure: 'One of the most widely used sample-based weighting methods is to follow up all or, more

6 Their approach should not be mixed up with Little and Rubin's (1987, p. 9) double sample approach which entails collecting information on all sample units and then extra information on a subsample.

usually, just a sub-sample of initial nonrespondents to a survey, and then weight the respondents to the follow-up to represent all the initial nonrespondents. The method is most commonly used in postal surveys where it is often difficult to obtain a satisfactory response.' He describes the procedure adopted in a postal survey among disabled people living in private households where the initial response was only 81%, which would today of course be highly satisfactory, even in a face-to-face survey. From the 19% refusers, a 10% sample was drawn which yielded an 82% response rate. Elliot considers this as evidence that there is no hard core of nonrespondents, or in the worst case only a very small one. Nevertheless, he is not very much in favour of the double sample approach. Firstly, he is concerned about including he respondents to the follow-up survey in the weighted data file, because the small sample fraction would reduce precision (and luckily differences in disability rates between first and second phase respondents were not significant). Furthermore, the additional costs of the follow-up interviews may be substantial, the method will usually involve some delay in completing the fieldwork, and 'there are the less tangible costs such as the risk of antagonising some of the sample members who are approached a second time for an interview having once refused. Thus the method should perhaps be limited to situations where nonresponse bias is a serious concern'.

Bethlehem and Kersten (1986, pp. 257-272) give an extensive, formal and statistical overview of Hansen and Hurvitz's double sampling approach, including variants, and show how it can be used to reduce error. They also suggest that a stratified second phase sample might perform even better, when stratification is according to reason for nonresponse (see the classes of nonparticipation model in the previous section). They assume that the nonresponse rate in the sample of refusers is twice as high as the nonresponse rate among the original sample (assumed to be higher than 50%). This makes their approach different from that of Hansen and Hurvitz, who assumed a 100% response rate in the second sample, and also from that of Groves, who stated that the double sample approach would still be useful if response rates in the follow-up sample were higher than in the original sample. They saw as drawbacks of double sampling that it is expensive and takes extra time, and leans heavily on the assumption that unbiased estimators can be computed for all nonrespondents (as the nonresponse in the second phase will presumably be high).

If nonresponse bias due to response rates below 80% is a serious concern, we should start taking the double sampling method seriously. Elliot, however, seems to be more taken with the central question method developed in the eighties by Bethlehem and Kersten (1986). This procedure implies that all nonrespondents are asked to answer one question central to the survey, either on the doorstep or by post or telephone, assuming a high response rate on this central question. In a labour force survey this could be, 'How many persons in this household have a paid job?', in a housing demand survey, 'Do you intend to move within two years?', etc. The same question should be asked in the regular survey, preferably as early as possible to prevent other,

more detailed questions influencing the replies. Bethlehem and Kersten also proposed a combination of both methods in which all nonrespondents, including those in the second sample, are asked the central question. The answer to the central question should be used to weight for nonresponse.

The central question method is relatively inexpensive and can in many cases be used when cooperation is being denied. Eliot feels this technique may be particularly useful when response is related to whether or not people see the survey as relevant to them. As Groves, Cialdini and Couper (1992, p. 481) have shown, preceding a modest request (one question) by a large request (long interview) might prompt refusers to cooperate. One problem is that the survey should have a 'central' question, and another, as observed by Elliot, is that the context of the central question may differ for those who answer it as part of a long questionnaire on this subject and for those who answer isolated question.

In the AVO1995 (an earlier round of the survey to be presented in chapter 6) the central question procedure was used rather unsuccessfully (Van Leest & Burhenne, 1997). Only about a third of the nonrespondents were willing to answer the short questionnaire. Among these respondents the elderly and nonworkers were overrepresented and the more highly educated underrepresented. It also proved to be very difficult to distil 'the' central question from a very long questionnaire in which a vast range of facilities, amenities and services play a role.

Lynn (2003) describes the PEDAKSI methodology, Pre-Emptive Doorstep Administration of Key Survey Items. This implies '... that the survey interviewer, having made contact with a sample member, should ask a small number of key survey items as soon as it becomes apparent that an interview is not going to be achieved at that visit to the address, even though it still might be possible to achieve the interview at a later visit. This is done by administering a Key Item Form (KIF)' (p. 241). In addition, Lynn used auxiliary data from the sampling frame, geographic data and interviewer observation data in his analysis. In a large-scale field test of PEDAKSI in the British Crime Survey, a survey with an 83% response rate, a KIF sheet containing 13 questions, six of them on crime, was completed by 6% of the final respondents and 25% of the nonrespondents. The information of the respondents was used to compare KIF outcomes with results of the questionnaire. KIF respondents were in some respects similar to survey respondents (entry phone, ethnic minority neighbourhood) and in other respects similar to nonrespondents (security devices, house condition, social status of neighbourhood). The nonrespondents who completed the KIF form appeared to have experienced fewer crimes than the respondents. Lynn ends with the suggestion that interview surveys for which estimates of nonresponse bias would be useful should consider adopting PEDAKSI methodology.

The central question procedure was used successfully by Voogt (2004), who recommends that each survey questionnaire should have three versions: the original one, a shorter one with essential questions and a central question(s)-only version. Voogt goes even further and posits that his central question ('did you vote at the last national elections?') is not only central in political studies and election surveys, but should be central in every survey as voting behaviour reflects social involvement and is therefore at the core of survey participation. Groves, Presser and Dipko (2004) report similar results as they found that, independent of the topic of the survey, political contributors participated more. Van Goor & Van Goor (2003) on the other hand, rejected the procedure as they obtained low response rates on their central questions, and also a higher rate of item nonresponse.

In the following chapters an important role will be played by a follow-up survey among a sample of persistent refusals. This survey was expected to provide information on a wide range of issues and from the majority of the refusals. The drawback of course is the necessarily small sample size of the follow-up survey (to keep costs within certain limits). This instrument to collect information on nonrespondents was preferred to the central question procedure for several reasons:

- in an earlier round of the survey the central question method was implemented rather unsuccessfully (low response rate, unequal response rate among different groups;
- in a multi-purpose survey it can be very difficult to identify the central question(s);
- in many field studies the response rate on the central questions was low and possibly unequal among different groups.

The choice of the follow-up approach does not imply that this is deemed to be better than the central question approach. One might argue with Bethlehem and Kersten that a combination of both approaches would be best.

5.5 *Conclusions from the literature: implementation of good practices and issues for further research*

In section 1.3 six central questions were formulated. Several questions have been the subject of extensive study the results of which have been implemented in fieldwork design and nonresponse analyses to be presented in the following chapters. The evidence on other questions is inconclusive, which is why specific research questions have been formulated that will also be addressed in the next questions. In section 2.6 the two different components of participation in face-to-face surveys were distinguished, i.e. being contacted and being cooperative. Throughout the empirical part of this study accessibility and amenability will be treated separately.

The subsequent empirical chapters use nonresponse data from two surveys, the Dutch Amenities and Services Utilisation Survey (AVO) 1999 to be introduced in chapter 6 and the European Social Survey (ESS) 2002/2003 to be introduced in chapter 10.

Most attention will be paid to the first survey, as this was supplemented by a large experimental nonresponse study. The ESS has been included to demonstrate that the nonresponse measurement tool used in the AVO can also be used in other studies, and to explore whether the AVO nonresponse results can be replicated in a survey on an entirely different topic. In addition, the ESS makes it possible to compare nonresponse results across countries and can give an indication of the extent to which the Netherlands is special.

5.5.1 *Why are high nonresponse rates a reason for concern?*

Section 2.1 showed that too much value should not be attached to response rates *per se*, primarily because there is an understandable though objectionable tendency to present these in the most favourable light. Nonetheless, for those involved in survey research nowadays the need to pull out all the stops to achieve a response rate of 60% is enough to cause sleepless nights. Sections 2.1 and 2.3 explained when and why nonresponse can have a detrimental effect on survey quality. Of course, nonresponse is no problem at all in the unlikely case where missing is completely at random (MCAR). When response rates differ across different groups in society (the missing at random, or MAR case) this is also not too serious, as long as reasons for nonresponse are not related to the topic of the survey. In the missing not at random (NMAR) case nonresponse cannot be ignored, because nonresponse is related to the topic of the survey. The summary of psychological, social and behavioural characteristics underlying survey nonresponse presented below, gives reason for concern. If response rates are related to social isolation, obeying the law, doing volunteer work, and being part of the mainstream culture, for instance, this would lead to biased results from surveys on social participation, criminal behaviour, voluntary work and cultural activities.

In summary, high response rates are difficult to achieve, and high nonresponse rates leave ample space for bias. Chapters 9 and 10 will explore to which extent the outcomes of the AVO and the ESS appear to be biased.

5.5.2 *Who are less likely to be contacted and to cooperate?*

Section 3.2 showed that contactability and urbanicity are negatively correlated: the population of urban areas is harder to contact and has a higher probability of remaining a noncontact. A simple reason for this is often ignored, namely that it may be more difficult to hire interviewers in big cities and those interviewers may be more reticent in making evening calls in certain neighbourhoods, and as a consequence less effort is expended on big-city dwellers (see Bethlehem and Schouten, 2003, discussed in section 2.2). Another reason for this negative correlation may be the different population composition in big cities (more single-person households, more ethnic minorities, lower incomes) and to practical impediments such as entry phones, which hamper access to the high-rise buildings that are more frequent in big cities. An additional impediment may be the higher rate of ex-directory households or mobile phone-only households in big cities.

Family size is positively correlated with contactability. Large families may be easier to contact simply because the likelihood that at least one family member is at home is greater. Families with small children are more likely to be housebound, women are more at home than men, older people more than younger people, and the unemployed, old-age pensioners and homemakers more than those who are in paid employment or follow a study.

The relationship between background variables and survey cooperation has been extensively studied and, although the same variables are mentioned again and again as indicative of reluctance to cooperate, closer meta studies have shown that correlations are often weak (section 3.3). Age, for instance, is often included in nonresponse analyses as an explanatory variable. Groves and Couper (1998) applied different analytic models and found less cooperation from the elderly, no age differences (after including household size), and a curvilinear effect of age in a multivariate analysis in which both young and old households cooperate more. When identifying age as a correlate of survey cooperation one should specify what age stands for: being part of a single-person household, having a lower education, being less adroit at filling in complicated forms, having a greater distrust of strangers or a higher sense of civic duty. These different causal factors may have a different impact on survey nonresponse dependent on design and topic, and will only partly surface through their relationship with age. Results for age are thus inconclusive, as are the results for gender. The presence of children in the household appears to be positively correlated with survey cooperation, as does household size in those cases where any responsible adult can participate. If all household members have to participate, household size is negatively correlated to survey cooperation (section 3.3.2).

There is some indication that persons with a lower educational level participate less in surveys, possibly because they have les experience with filling in complicated forms (which would be especially relevant in self-completion surveys) and may see surveys less as an interesting intellectual activity (section 3.3.3). The results on socio-economic status, a characteristic related to education, are again fairly inconclusive, although most evidence seems to suggest that cooperation is higher in the higher social strata.

Urbanicity is negatively correlated with both contactability and cooperation and in both cases the relationship is complicated. Lower cooperation rates in urban areas (section 3.3.4) may be a consequence of differences in population composition between rural and urban areas, or of less cohesion and less trust in strangers in big cities. Ethnic minorities in the Netherlands frequently live in big cities, which will make urbanicity and ethnic origin and possibly even religion (Islam) difficult to isolate as separate factors. The presence of multiple factors behind urbanicity as a correlate of noncooperation shows why oversampling urban areas to compensate for low response rates (see section 2.4.2) may not reduce nonresponse bias. If refusal in big cities is high because of the high refusal rate of (for example) students who pre-

dominantly live in big cities, including additional non-student big-city dwellers will not solve the problem at all.

The main conclusion of chapter 3 was that background variables should be treated as correlates of nonresponse and not as causes. When looking under the surface, a number of psychological, social and behavioural characteristics were identified in chapters 3 and 4 that were more likely causes of nonresponse or that at least reflected these causes: social isolation, social participation, abiding the law, interest in societal well-being, doing volunteer work, political interest and knowledge, political cynicism, electoral participation, form literacy, lifestyle (section 5.3), being part of the mainstream culture, type of leisure activities, PC ownership and use, Internet use, the type of sponsor, and attitudes towards surveys (survey value and survey enjoyment). In addition, there is clear evidence that survey cooperation depends on interest or involvement in the topic of the survey. There is no conclusive evidence that being busy (section 3.3.5) is a reason not to spend time on surveys; busy active people might even just take a survey in their stride.

The results from the literature on who is less likely to participate have been implemented in the design of the follow-up survey among persistent refusals in the AVO (see section 6.3). For this survey three groups of variables were selected from the large AVO questionnaire: background variables including age, urbanicity, dwelling type, neighbourhood characteristics, family type, age, gender, daily occupation, education and income. In addition a subset of core variables from the AVO was selected. In selecting this subset variables related to social participation, leisure activities, membership organization, etc. were given precedence as these appear to be related to the causal factors above.

The relationship between survey nonresponse and background variables is not the topic of this study. It has been well covered in the literature and it is not to be expected that a single survey will put things in an entirely different light. What this study can add stems from the fact that contact (chapter 7) and cooperation (chapter 8) will be analysed separately.

5.5.3 *How can response rates be enhanced?*

According to section 1.1 the survey climate in the Netherlands is not very favourable to obtaining high response rates. The Netherlands appears to be the lame duck in survey research and is looked down upon by other countries that generally achieve better results. The high nonresponse rates in the Netherlands seem to be one of the reasons why Dutch market research organizations are increasingly setting up access panels in which high response rates from cooperative panel members are guaranteed (and initial low response rates are ignored, see section 2.4). Random sample surveys, however, are still the preferred research instrument of most statistical offices and academic and governmental research organizations. High target

response rates are inextricably linked with random sample surveys, and much harder to achieve than in non-probability access panels.

Response rates in face-to-face surveys can be enhanced by increasing the contact rate and the cooperation rate (see section 2.2). In theory, maximizing the contact rate is simple and according to rules of thumb not yet contacted sample units should:
- be contacted at different times of the day and on different days of the week to deal with standard patterns (working hours, sports clubs);
- receive a certain number of calls to maximize the probability that they will be found at home at least once;
- be called on when people are most likely to be at home, and not only during (interviewer) working hours; and
- be called on after an intermission of some time in case of a prolonged stay elsewhere (hospital, abroad).

Furthermore, practical impediments should be overcome (security devices, vicious dogs, entry phones, doormen). Under certain conditions respondents can be recruited by telephone, even in face-to-face surveys. Information from neighbours ('they will be back next week') or from interviewer observations (child's bike in the garden) can guide the interviewer in optimising the timing of future calls after unsuccessful earlier ones.

The strategies for enhancing contactability that were implemented in the AVO will be presented in chapter 6. Chapter 7 will describe the field efforts it took to achieve a high contact rate and show how call efficiency can be improved.

When prospective respondents have been contacted they will be asked to cooperate. Section 4.3 presented an overview of the literature on obtaining cooperation in which survey preparation and doorstep interaction were highlighted. A first major factor is hiring, training and motivating experienced interviewers who should have the means to tailor to the preferences and needs of respondents. A second major factor is that respondents differ. Knowing which groups in society may be less likely to cooperate may give some guidance in how to obtain cooperation. Reasons to cooperate may vary from interest in the topic, perceived importance of the survey sponsor, liking the interviewer and being a law-abiding citizen to receiving a monetary incentive. In the process of obtaining cooperation the interviewer should cater as much as possible to individual triggers. This can be helped by information on reasons for refusal given at earlier contacts. Section 4.2.2 distinguished situational factors for non-compliance (too busy, napping, entertaining) which may be more or less random factors and not produce nonresponse bias, and survey-related factors (distrust of strangers, negative attitudes towards surveys, disinterest in the topic) which might have a more systematic and devastating effect. When trying to enhance response rates both types of reasons for refusals should be taken into account.

Section 4.3 gave an overview of costs and benefits of being interviewed, based on the assumption that in order to enhance response rates the perceived costs of being interviewed should be minimized and the perceived benefits maximized. Since reasons for survey cooperation and noncooperation differ, tailoring is of paramount importance in interview surveys, i.e. the interviewing emphasizing those survey characteristics that are salient to the respondent, insisting when there is an opening and temporarily withdrawing before eliciting a final refusal. Tailoring is more than playing lip service to the respondent's perceived interests and preferences. If possible, an interview should be conducted when, where and how it suits the respondent. The interviewer should be willing to come back at a more convenient time and if necessary and possible conduct the interview at a public location or by telephone or leave a self-completion questionnaire. This will not always be possible, and introducing alternative interview modes may introduce alternative survey errors (see section 2.4.3) but it is likely to enhance response rates.

A general requirement in both establishing contact and achieving cooperation is a sufficiently long duration of the fieldwork and the deployment of a sufficient number of interviewers in order to make a sufficient number of calls on people who are rarely at home, to catch people at home who have been away for a time, to substitute a poorly performing interviewer by a more successful and motivated one, to re-issue an interview and to try refusal conversion by a new interviewer. This requires close monitoring of the fieldwork and means that achieving a high response rate may take a while, which may make a survey less timely, whereas timeliness is another quality criterion.

The accumulated knowledge on obtaining cooperation has been implemented in the design of the AVO and the follow-up survey among persistent refusals that will be presented in chapter 6.[7] This chapter will set out how cooperation was obtained from extremely resistant respondents (section 6.3) and explore why the follow-up survey succeeded. Guidelines on how to achieve high response rates in the ESS 2002/2003 were given in a document drafted by Achim Koch from ZUMA (ESS, 2004). Response enhancement in the ESS will be discussed in chapter 10.

Obtaining the cooperation of initially reluctant respondents, also called refusal conversion, was described in section 2.6.3. Refusal conversion was an essential part of the fieldwork design of the AVO (section 6.2). Section 8.2 will show how refusal conversion contributed to the response rates of the AVO, and section 8.3 will explore which groups are especially susceptible to refusal conversion, and if the success of refusal conversion attempts is related to the initial reason for nonresponse.

7 In preparation to the follow-up survey, Frans Louwen, the research director of GfK, the survey organization that collected the data, and the author attended the International Conference on Survey Nonresponse, held in Portland, Oregon in October 1999. They presented their plans to study nonresponse (Stoop and Louwen, 1999) and, more importantly, learned form survey experts.

The final section of chapter 4 discussed response enhancement from a moral point of view and pitted behavioural theories in which sample cases are manipulated and cajoled into saying yes to the interviewer request against the social exchange view, in which responsible citizens are free to decide whether or not they want to cooperate in a survey. Are our efforts at response enhancement ethically defensible? The present study adheres to the principle that social surveys for academic and policy purposes work for the general good and need high response rates, and that it therefore seems acceptable and wise to tailor the survey approach to the many different reasons a respondent may have to cooperate, while keeping ethical guidelines in mind. A related issue is whether putting pressure on respondents to cooperate has a detrimental effect on data quality. There is no evidence that positive pressure, i.e. handing out incentives, harms data quality (section 4.3.2). Voogt (2004, section 5.3) noticed that repeatedly asking respondents to cooperate might lead to irritation and satisficing behaviour. To find out if converted refusals resort to satisficing, it will be explored whether converted refusals deliver data of poorer quality (section 8.4 and 9.3.1).

5.5.4 *How to study nonresponse?*

Section 5.1 gave an overview of approaches to studying nonresponse. The present study combines a number of approaches in order to find out if extended field efforts and, hopefully, enhanced response rates, reduce bias. This would occur if through extended field efforts increasingly difficult respondents, or a particular kind of respondents would participate. This is essentially the idea of the two models of Lin and Schaeffer (1995, see section 5.2) for assessing the impact of nonparticipation on survey estimates, namely the continuum of resistance model and the classes of nonparticipants model. The first model assumes that there is an underlying continuum on which all potential participants are ordered and that deploying more efforts results in obtaining the participation of more participants. The second, more qualitative model distinguishes between classes of nonparticipants. Specific classes of participants, e.g. temporary refusals and hard-to-contacts, can be used to estimate the impact of different types of nonparticipation, e.g. final refusals and noncontacts. Whereas the first model assumes that nonparticipation is a one-dimensional phenomenon and that obtaining the participation of the most resistant sample persons will reduce bias, the second model considers nonresponse as a multi-dimensional phenomenon and assumes that different groups of participants can be used as proxies for different kinds of nonparticipants. The literature presented in section 5.2 shows that the evidence on both models is inconclusive and that they are rarely tested stringently. What is generally done is either to test a continuum of resistance with respect to contactability, or compare three classes of participants (early, late and reluctant), or compare different reasons for refusal. What is generally unavailable is information on final nonparticipants, which would be necessary to fully test the models.

The present study aims at rigorously testing variants of these models. The single continuum of resistance model does not seem to be very useful as contactability and cooperation are clearly different factors in survey cooperation.[8] Therefore, the existence of a continuum of resistance with regard to both contactability (section 7.3) and cooperation (sections 8.1 and 8.3.1 for the AVO, and section 10.5 for the ESS) will be explored. The classic 'classes of nonparticipants' model in which easy, late and reluctant respondents are compared will be presented in section 8.3.2 (without final refusals) and section 9.3.2 (with final refusals). The test of the latter model will show if temporary refusals can be used as proxies for final refusals. In addition, classes of participants will be distinguished according to reasons for refusal and interviewer assessments of future success (AVO, section 8.3.2 and ESS, section 10.5).

To test these models information on contactability and reluctance of respondents is required as well as information on final nonrespondents. As set out in section 2.6.4 paradata, which capture the individual progress of fieldwork and make it possible to chart contactability and reluctance to cooperate, are generally deemed of great importance and rarely collected. Section 6.2.3 will present the AVO 1999 contact form, the instrument with which this information can be collected, show to what extent they were satisfactory and how they could be improved. Similar contact forms were used in the ESS 2002/2003 the development of which will be described in section 10.3.[9] Chapters 7, 8 and 10 will show that the information from the contact forms make it possible to check whether fieldwork has been carried out according to the specifications, to explore how difficult it was to reach sample units and to persuade contacted persons to cooperate, to investigate which refusals were reapproached and which were converted, to study bias due to different reasons for refusal and to calculate interviewer effects. These chapters will also show that even with the help of detailed, well kept contact forms, measuring contactability and reluctance to cooperate is far from simple. Section 7.2.1 will outline why it is difficult to measure contactability and sections 8.1 (AVO) and 10.5 and 10.6 (ESS) reluctance to cooperate. The main reason why this is difficult is because contactability and reluctance to cooperate are not innate characteristics of respondents but are an outcome of fieldwork design, fieldwork decisions and interviewer behaviour.

In addition to information on contactability and cooperation, the models to be tested require that information is available on the final nonrespondents that are either at the end of the continuum of resistance, or that comprise different classes of nonparticipants. Besides collecting paradata on respondents and nonrespondents, interviewer

8 Although the difference between contactability and cooperation is one of the underlying themes of this study, empirical evidence for this distinction, based on the contact forms, will be presented in section 7.4.

9 Based on the experiences with these contact forms improved ones have been developed and used in respectively the AVO 2003 and the ESS 2004/2005.

observations of sample households, the dwelling and the neighbourhood, frame data and geographical database, two instruments for collecting independent information on nonrespondents have been studied, namely the Central Question Procedure introduced by Bethlehem and Kersten (1986) and a follow-up survey among a sample of persistent refusals. As set out at the end of section 5.4, in this study the second approach will be implemented. Section 6.3 will describe the design of the follow-up survey and section 9.2 will demonstrate which practical problems had to be overcome in order to obtain information on persistent refusals that can be compared with the regular survey.

5.5.5 *Do enhanced response rates improve the accuracy of survey results?*

It is generally assumed that extended field efforts result in higher response rates, which in turn produce more accurate survey estimates. To find out if this is true one can compare low-effort, low-response surveys with high-effort, high-response surveys and assess whether the latter give more accurate estimates of a variable whose true value is known. This was done by Merkle and Edelman (2002, see section 5.3) who found in a study of exit polls that an increase in response rates may even increase bias. Other studies on the effect of extended field efforts presented in section 5.3 showed that high or higher response surveys are not necessarily better, and that the added value of increasing response rates when they are around 70% seems questionable. The literature furthermore indicates that those who are hard to contact differ from the easily contactable and that this difference is larger than that between immediately cooperative and initially reluctant respondents. Finally, temporary refusals are not necessarily similar to final refusals.

In the present study it will be possible to assess whether nonresponse bias is reduced when enhancing response rates, as information on final refusal is available and those who need extended field efforts can be identified. With the models and instruments announced in the previous sections, the question of whether enhanced response rates improve the accuracy of survey results can be addressed in the next chapters. Chapter 7 will explore whether a higher contact rate improves accuracy; chapter 8 will analyse the yields of refusal conversion, and chapter 9 will assess whether it is likely that enhancing the contact and cooperation rates has improved the accuracy of survey outcomes. Section 10.5 will present results on response enhancement in the ESS, although no information on final refusals is available for this survey.

5.5.6 *How to combat nonresponse error and allocate funds effectively?*

Merkle and Edelman (2002) concluded from their study on nonresponse bias that: 'This should serve as an eye-opener to those who pursue higher response rates for their own sake without considering the impact of survey error. Devoting limited resources to increasing response rates with little or no impact on survey error is nor money well spent, especially when that money might be better spent reducing other

sources of error...' Similar warnings have been included in this study from Groves (1989), Krosnick (1999), Platek and Särndal (2001) and Biemer and Lyberg (2003). This study is on nonresponse error rather than on total survey error. Therefore, the final question will be how to allocate funds to either response enhancement or nonresponse measurement effectively in order to reduce nonresponse bias. This is in line with Bradburn's (1992) recommended approach to the nonresponse problem, namely '... to allocate the data collection resources in a more targeted manner to learn more about the possible bias arising from low response rates' (see section 5.4).

The question raised in the heading of this section will be addressed in chapter 11, after reviewing the empirical outcomes of this nonresponse study presented in chapters 6 to 10.

Part 2

Nonresponse in face-to-face surveys: an emperical study

6 Going into the field

6.1 Introduction

This chapter describes the fieldwork for the Amenities and Services Utilization Survey (*Aanvullend Voorzieningengebruikonderzoek* or AVO in Dutch) carried out in 1999/2000. This fieldwork has been meticulously documented and extensive paradata have been collected and keyed to facilitate a detailed analysis of the response process. In addition, the main survey was followed by a survey among a sample of persistent refusers, set up to compare respondents and nonrespondents. Section 6.2 describes the fieldwork of the main survey, the paradata that were collected and the problems encountered in collecting data and paradata, and ends with suggestions for the improvement of fieldwork and paradata collection. Section 6.3 describes the design, the fieldwork and the evaluation of the follow-up survey. Interviewer records and focus group discussions show why and how a very high response rate was obtained. This section is based on Stoop (2004a) *Surveying Nonrespondents*, published in *Field Methods*.

6.2 Amenities and Services Utilization Survey 1999 (AVO)

6.2.1 History and content

Since 1979 the SCP has commissioned the data collection for the four-yearly Amenities and Services Utilization Survey. The data are used by the SCP, Statistics Netherlands (CBS) and a wide range of academic, policy and commercial researchers[1]. The AVO comprises a household questionnaire and a drop-off questionnaire for every household member aged 6 years and older. It covers a wide range of topics. One set of questions measures general socio-structural characteristics (household composition, characteristics of the dwelling, employment, education and income), other question blocks cover the utilization of a wide range of generally publicly-funded services, facilities and amenities. An overview of the 1999 questionnaire is given in table 6.1. The content of the questionnaire varies to some extent between rounds, but the major part of the questions remains the same.

In 1991 the response rate in the AVO reached a meagre 43%, as opposed to approximately 60% in earlier years (see table 6.2). In the public procurement procedure for the AVO1995 a high response rate was set as a major selection criterion. The organization that had been selected to gather the data (GfK Panel Services Benelux)

1 The data files are available via the Scientific Statistical Agency of the Dutch Science Foundation (NWO) or the Steinmetz Archive.

guaranteed a response rate of 70%. This was higher than usual in the Netherlands and, surprisingly, according to them even less expensive than a 60% response rate as it was less expensive to re-approach initially selected households than to draw a large gross sample and approach new potential respondents to obtain the required net sample. The high target response rate was reached through a readjusted survey design, setting a high limit for the number of calls, converting initial refusers, deploying better educated, accredited interviewers and paying special attention to the interaction with sample households.

In 1995, the impact on response of a large number of calls and converting refusals could not be distinguished, as only information on the total number of calls was available and not on the number of calls to first contact and temporary refusals. The response rate would have been 50% in 1995, however, if the maximum number of calls had been restricted to three, as in 1991 when one of every eight households did not cooperate because the interviewer did not find them at home on one of the three calls allotted to each household. In 1995 this percentage was negligible. The percentage of final refusals, however, stayed more or less the same, at 25.6% in 1991 compared with 23.6% in 1995. The latter percentage would have been higher without active efforts to achieve refusal conversion.

Table 6.1a Overview of questions in the VO1999: household questionnaire (face-to-face)

subject	questions
family composition (per member)	year of birth, sex, marital status and position in household
dwelling	type, floor, lift available, size, age building, duration occupation, extra facilities
dwelling costs	owned/rented, value, rent, mortgage information, rent subsidy
environmental behaviour:	separation of waste, information on car(s)
care received during last 12 months	
personal care	from whom, type, frequency, period, why, infant welfare centre, medical appliances
domestic care	from whom, period, amount, reason, costs, homecare
for those who receive home care	use transport and other facilities for the handicapped
child care and education	
number of children	
seeking help, pedagogic problems	
children 0-4 years (per child) parental leave, (in)formal day care per child (type, cost,	parental leave, (in)formal day care per child (type, cost, frequency), kindergarten, day care facilities
children 4-12 years (per child)	type of school per child, child care after or in-between school hours, costs
children 12-30 years (per child)	living at home, student grants, child benefit, education (type)
income	number of earners, family income
informal labour	for other households or organizations (participation, remuneration)

Table 6.1b Overview of questions in the VO1999: individual questionnaire (paper and pencil)

subject	questions 16+ years	4 pages	questions 5-16 years	4 pages
sports activities		5		3
use recreational facilities		3		3
cultural participation	theatre, concerts, pop music, ballet, movies, art show	4	theatre, concerts, pop music, ballet, movies, art show	4
media use	reading, library, computer, radio, TV, CD, tapes, VCR	3	reading, library, computer, radio, TV, CD, tapes, VCR	3
arts and crafts	drawing, sculpting, singing, theatre, photography	2	drawing, sculpting, singing, theatre, photography	2
social network		2		
present education		4		1
educational level		2		
education and labour market position parents		2		
use counselling (labour market and other)		2		
miscellaneous	travel cards, membership organizations	1	children's helpline travel cards, membership organizations	1
contacts with police and legal aid		4		
medical care and health insurance		2		1
health problems	visual, hearing problems, problems with daily and household activities, health problems and ailments	6	visual, hearing problems, problems with daily activities, health problems and ailments	4
care leave		2		
religion, ethnicity		2		2
occupation		4		
disablement		2		
income	wage, benefit, pension, etc.	7		

6.2.2 *Fieldwork design, implementation and outcomes*

Most face-to-face surveys in the Netherlands use either the population register as a sampling frame or the Postal Addresses File (PAF). The population register is obviously the best frame to draw a sample from, especially as it contains additional information on each member. One drawback, however, is the limited access for survey research and the potential errors due to the absence of the homeless and illegal inhabitants and – more seriously – the incorrect addresses of students and subtenants. In 2003, for instance, according to the Dutch daily *De Volkskrant* of 1 April of that year, one in eight first-year students did not live at their official address, either because they lived with their parents but had their official address elsewhere

Table 6.2 Response rates in the VO survey

year	net sample (eligible addresses)	responding households	responding persons	response rate[a]
1979	9, 915	6, 431	17, 232	65
1983	9, 908	5, 774	14, 869	58
1987	10, 302	6, 496	16, 151	63
1991	12, 797	5, 458	13, 105	43
1995	9, 305	6, 421	14, 489	70
1999	9, 372	6, 125	13, 490	65

a Methods of computing response rates may differ per round

– in order to obtain a higher grant – or could not register at their actual residence because their landlord did not want it known that he rented out rooms. The PAF sampling frame, comprising all (residential) addresses in the Netherlands, is less prone to contamination than the population registers: everybody who lives somewhere will in principle be found. The drawback of this frame is that it will not always be clear whether an address is residential or not, people may live at two addresses (one being a holiday home, or because they have an additional residence close to work) and the enumeration of households at addresses and persons within households can only take place after an address has been sampled and after a household has been contacted (see section 2.6.2).

The sampling and fieldwork approach of the AVO1999 is depicted in figure 2.2. The implementation will be discussed here and problems with the design in the next section. The *population* of the AVO comprised all households living in the Netherlands at residential addresses. As a *sample frame* the Postal Addresses File was used, from which business and institutional addresses and post office boxes were excluded, so that residential addresses (including farms and houseboats), shops and homes for the elderly remained. The latter two categories are bound to include ineligibles but also a number of residents who share their letterbox with a shop (among whom the author) or a resident member of the staff of a senior housing facility. From the sampling frame a *gross sample* of about 10,500 postal addresses was drawn. By linking these (electronically) to the telephone directory, for 75% of the addresses the telephone number and the name of the inhabitants could be found. The telephone number could be used by the interviewer in later stages of the fieldwork. The absence of a listed phone number in the telephone directory could indicate an ineligible sample unit (empty dwelling), no landline telephone or an unlisted telephone number. By adding a name to the address, personalized advance letters could be sent. These advance letters and a brochure providing details of the survey and the fieldwork, the survey organization and the sponsor were mailed shortly before the first contact attempt to each address, either by name or, if no telephone number – and no name – was available, 'to the inhabitants of ...'.

This personalized approach has clear advantages, but also two drawbacks. Firstly, if more than one household resides at a given address, only one receives an advance letter. This can be solved by sending another advance letter to the other households at this address that are selected when the interviewer calls. The second problem is that the individual addressed may have moved or died. Theoretically, this is not a problem, because the sample frame comprises those persons living at the address when the interviewer calls. If the advance letter is returned to sender by the post office, indicating a change of address or death, a non-personalized letter can be sent and the interviewer can simply visit the address and ask the present residents to cooperate. The situation becomes more complicated when the individuals living at the address themselves return the advance letter, which happens quite regularly. In one case they even sent the copy of the death certificate of the (deceased) person to whom the advance letter had been addressed. In this case the interviewer was to visit the household and explain that the survivors can participate instead, which is theoretically sound but may be somewhat awkward in practice. In these cases ineligibility might be a fuzzy category.

The fieldwork was subdivided into three identical waves. At the first call[2] of the first wave interviewers had to complete a short list of questions on the dwelling and the neighbourhood of each sample address (annex 6.5). They also had to assess whether the address was eligible or not. Ineligibility may be obvious from the start, for instance when a building has been demolished. In other cases, contact with neighbours on the first or subsequent visits may provide the necessary information for classifying an address as ineligible. Finally, for some addresses it may remain uncertain whether the household at the address was never at home when the interviewer called or the dwelling was empty. Besides having to assess the eligibility, the interviewer had to determine at the first call – or at the first contact – the number of households at the address. If an address comprised more than one household, up to four households per address had to be interviewed. For each additional household a separate contact form had to be filled in. The selection of the households at addresses where more than four households lived was left to the interviewer (see annex 6.1). After the first wave all questionnaires were returned to the interview organization. Noncontacts and temporary refusals were reissued for a second wave of fieldwork, and unsuccessful cases from the second wave were reissued for a third wave. At the end of each wave, the interviewer had to assess for as yet unsuccessful

2 In accordance with the nonresponse literature the term *call* will be used throughout this study for a personal visit to a sample address or household. A call is a *contact attempt* and may or may not result in a *contact*. Once a call has resulted in a contact, the interviewer shifts from establishing contact to obtaining cooperation. The number of calls to first contact is a measure of contactability (see chapter 6) despite the fact that after the first contact there may be subsequent calls without contact. The interviewer may also make *telephone calls* to the household in later stages of the fieldwork. These will be identified as such, if possible.

cases whether success in future waves was likely and whether different interview modes might be more successful.

In each wave, the interviewers had to make at least three personal calls to the selected addresses at different times of the day and on different days of the week until first contact was established. Each call had to be registered on a contact form for each household (see annex 6.4). After three unsuccessful calls the interviewer was to leave a 'not-at-home' card stating that the interviewer had called when the respondent was not at home. The card gave the name and phone number of the interviewer, and invited the respondents to get in touch with the interviewer and make an appointment. Alternatively, the interviewer could try to make contact by telephone in order to complete an interview. Telephone contact attempts initiated by the interviewer were thus only allowed after three calls at which the respondent was not at home.

All parties involved – sponsor, survey organization and interviewers – were rather reticent about making telephone calls to the household earlier in the fieldwork process, despite the fact that this could make unsuccessful contact attempts superfluous, as they feared respondents more easily refuse their cooperation on the telephone (see section 2.5). After three futile contact attempts, the balance between an increased probability of refusal and not contacting the household at all dipped to the first alternative. If a telephone appointment had been made, either on the initiative of the respondent or of the interviewer, the interviewer visited the respondent to conduct an interview.

The contact forms of the addresses that had not been completed in the first wave, either by an interview or by declaring them ineligible, were inspected by the fieldwork organization. Addresses that had not been reached, and addresses at which respondents had refused or said they were not able to cooperate but where a new approach still seemed viable, were reissued. The decision to reissue a refusal was based on the evaluation of the fieldwork organization of the interviewer ('very persistent', 'may give up too easily', 'had a temporary low point'), the reason for refusal as noted down by the interviewer and the assessment of the interviewer of possible future success. The reissued interviews were assigned to interviewers, either the same as in the first wave or a new one and a second, similar wave began. Most interviewers were deployed in all three waves. The decision to select a new interviewer or keep with the earlier one was partly based on logistic and economic reasons, and partly on the assessment of the fieldwork organization of the likelihood of the interviewer to succeed in the new wave. This second wave was succeeded by a third one, if necessary, according to a similar process of re-issuing. Annex 6.2 gives an indication of the distribution of fieldwork across the entire fieldwork period, and in particular across waves.

Table 6.3 presents the outcome codes for the AVO1999. The hard figures disguise minor uncertainties. A close study of the paradata confirms how difficult it is to distinguish between ineligible and noncontacted addresses or uncooperative respondents. Firstly, it turned out that 320 sample units where no contact had been established were classified as ineligible. This was the result of the standard procedure in the AVO1999 where the survey organization classified all noncontacted units after nine repeated calls as vacant and thus ineligible (GfK, 2000). Closer inspection of the call records indicates that in 172 cases the address was indeed an empty dwelling or a non-residential address. In the remaining 148 cases all 9, 10, 11 or 12 contact attempts were futile but there was no independent information that these were empty dwellings. The noncontacts that were classified so as a result of this standard procedure have been reclassified here as noncontacts, and thus as nonresponse. Due to this reclassification, an unknown number of ineligible cases will erroneously be classified as nonresponse.

Secondly, in the AVO households in which no adult spoke Dutch were to be excluded, and thus ineligible, according to figure 2.2. Classifying persons who do not speak the fielding language is contestable.[3] Had the 133 households in table that were excluded because of language problems been classified as nonresponse, the non-response rate would have been 64.2% instead of 65.1%. Of course, the presence of a language problem can only be determined after contact. As a consequence, a number of the noncontacts might be households that would be excluded, had they been contacted, because no member of the household spoke Dutch. In addition to the 133 'language' households, for 24 households language was also recorded as a problem. On other calls at these households, however, it seemed that the language problem was more likely an excuse for refusal. At subsequent calls, five of these 'not able because of language problems' resulted in an interview and a completed set of questionnaires. The presence of language problems can thus only be asserted after contact, and the assessment of language problems (exclusion) may hide noncompliance (nonresponse).

3 The AVO follows here the approach described by Groves (1989, pp. 137-138): *The tradition in some organizations is to attempt no interviews with persons who cannot speak English (i.e., they are removed from the denominator of the response rate calculations and from the target population).* In a footnote he gives as a possible rationale for this tradition ... *concerns both with nonresponse error and measurement error associated with this population. Even when this group might be willing to respond, the translation of survey questions into another language and of the respondent's answers to English might be productive of errors in the recorded data.* In the European Social Survey (see chapter 9) noncooperation due to language problems is classified as nonresponse, although the study specifications outlined that translation of the questionnaire was only required for minority languages that were spoken by more than 5% of the population of a country.

Table 6.3 Fieldwork outcomes V01999

fieldwork outcome	n		single	multi	total
gross sample (T)	10,623	addresses	9,173	49	9,222
lost , not used	249	households according to contact form	9,173	121	9,294
erroneous address	699	households accounted for	9,173	88	9,261
uninhabited	172	addresses with at least one responding household (R)	6,057	43	6,100
language problems	133	responding households	6,057	68	6,125
total ineligible	1,251				
net sample (N)	9,372	outcomes per address	%		
never at home	148	eligible addresses (N/T)	88.2		
contact (C)	9,222	contact rate (C/N)	98.4		
single household addresses	9,173	cooperation rate (R-total/C)	66.1		
multiple household addresses	49	response rate (R-total/N)	65.1		

6.2.3 *Paradata, observational data and frame data*

In each wave of the AVO a contact form had to be filled in on which each call, the reason for refusal and the assessment of future success had to be recorded (see annex 6.4). At the first visit the interviewer also had to complete a short list containing questions on the dwelling and the neighbourhood (annex 6.5). For each call a unique interviewer identification number was available, plus age and sex of the interviewer. Besides paradata from the contact forms, observational data from the interviewer, and interviewer numbers and background, two databases that could be linked to the sampling frame were available. The GeoMarktprofiel (GMP) Database contains information on 6-digit level postcodes (approximately 16 addresses). Information is available from inhabitants of the area (in a period of 4 years two inhabitants of every postcode are asked to provide information on their surroundings by telephone or mail) and a large number of registers (see www.wegenerdm.nl) The WoonMilieu-Database (WMD) comprises information on 4-digit level postcodes (several thousand addresses) from GMP, Statistics Netherlands, administrative records and public registers. A final variable of interest is the availability of a listed phone number for the sample addresses. In the AVO the presence of a listed phone number could influence response rates in two stages of the sample and contact procedure. If the name of the resident (from the phone directory) was known, advance letters could be sent addressed in person, instead of 'to the inhabitants of ...' (see section 4.3.3). Secondly, after three unsuccessful calls, the interviewer could make a phone call to the sample household and try to make an appointment. As a consequence, the presence of a listed phone number will have influenced fieldwork strategies. Besides, the presence of a listed phone number may be related to socio-demographic and maybe even psychological characteristics of the sample household members.

In analysing the data it turned out that that even the detailed paradata that were collected in the AVO showed some deficiencies in addition to those mentioned in the previous section; on the other hand, inspection of the paradata revealed some deficiencies in the fieldwork. Firstly, the paradata showed that multiple household addresses had not been treated as meticulously as they should have been. In the AVO no rules were given for household selection at addresses and the contact forms do not make entirely clear the number of households that should have been interviewed at each address. It is highly probable that not all households at multi-household addresses that should have been part of the sample were approached (see annex 6.1). Therefore the final response rates may be slightly overrated and single-person households in student apartments, for example, may be underrepresented. The effect will be small, however, as only about 3 per cent of Dutch households lived at multi-household addresses. As the number of addresses with more than four households is very small, the bias caused by non-random selection of households at addresses will be very small.

Secondly, the paradata proved to be incomplete with respect to interviewer identification. Annex 6.3 shows how interviewers were assigned across waves. Here a gap becomes visible: for about 20 percent of the gross sample no interviewer ID is available, nor the age and sex of interviewer. These calls with 'interviewer unknown' were made by a subcontractor working mainly in the western, most densely populated area of the Netherlands. This means that interviewer information will probably be biased and no analysis of interviewer effects will be presented. Since the majority of the AVO questions were self-completion and fairly factual, no interviewer effects on answers are expected. Interviewer effects on response behaviour could have occurred but cannot be analysed. These effects will have been mitigated by the deployment of another interviewer in a subsequent wave.

For the majority of cases where interviewers were identifiable, assignment sizes and deployment over waves have been computed (see annex 6.3 for a complete overview). Interviewer assignments in the AVO were fairly large. In the first wave, interviewers were assigned approximately 39 cases each. Across waves, interviewers were assigned approximately 66 cases. There are large differences in interviewer assignments. Most interviewers, about 81%, were deployed in all three waves, 6% only in the first wave(s) and 13% only in later wave(s).

Furthermore, it turned out that the timing of calls could not be reconstructed completely from the paradata.. Calls at the sample addresses were to be spread over the times of the day and the days of the week. On the contact form the time of day (morning, afternoon, evening), the day of the week (Monday to Saturday) and the month were recorded. This choice of rather large time slots made it impossible to distinguish between early and late afternoon calls and to assess the field outcomes from contacts around dinnertime. The coding of days of the week instead of days made it impossible to determine the influence of the time lapse between two calls.

A call on Wednesday morning and a subsequent one on Wednesday afternoon in the same month may be four weeks apart, but also a few hours only. With hindsight, it would have been preferable to have interviewers simply record the time of the day in hours and the date.

As a consequence of this deficient recording, the effect of the time lag between calls could not be analysed, nor was it possible to discover when a subsequent call should be made dependent on the timing of the previous call. Campanelli, Sturgis and Purdon (1997, p. 3-11) could and did study the probability of contact at the second call by the length of time since the previous call (in days). Their estimates suggest that the highest probabilities of contact are associated with either calling on the same day as the first call or after an interval of two weeks or more. They themselves contest the validity of these outcomes, however, as the second call on the same day will more likely be an evening call, and a second call after two weeks may indicate that the interviewer had obtained information that the household was away for the period.

The contact forms were also not designed to record telephone calls. At the end of each wave after three visits to the sample household's address, the interviewer could make a telephone call or be called by the sample household in response to a 'not at home' card left by the interviewer. The outcome of the telephone call could be recorded in the column for the fourth call (see annex 6.4). In theory either of the following could have occurred:

- If the interviewer did not know the phone number of the respondent, he or she may have paid a fourth visit and recorded the outcome in the fourth column (or paid even more visits and recorded the outcome of the last one in the fourth column).
- If the interviewer did know the phone number of the respondent, he or she still may have paid a fourth visit and recorded the outcome in the fourth column.
- The interviewer rang up the household once or more but never managed to find the household at home. In this case 'not at home' may have been recorded in the fourth column.
- The interviewer may have had telephone contact with the sample household (because he or she rang up, or because the interviewee called the phone number on the card) at which occasion the household representative told that he or she was not able or not willing to be interviewed. In this case, the outcome of the phone conversation will very likely have been recorded in the fourth column as refusal/not able.
- The phone contact between interviewer and sample household resulted in an appointment. As the interviewer got paid for an interview, it is a fair guess that the interviewer himself or herself conducted the subsequent interview at the appointed time. In this case the appointment will not have been recorded, only the outcome of the appointment. (The data seem to confirm this as, contrary with other calls, the outcome of the fourth call in the first wave and the eighth call in the second wave is never an appointment).

Consequently, it is unknown whether what went on during the fourth call was settled in a telephone contact or not. The contact form comprises one additional question pertaining to telephone contact per wave: 'If contact was made by telephone: How many times did you call before an appointment was arranged?'. With hindsight this seems to be a prototypically bad question as it refers only to telephone approaches made by the interviewer, only to telephone calls that result in contact and only to telephone contacts that result in an appointment. One wonders what kind of telephone attempts or contacts have been recorded by the interviewer and which have not. What is clear, however, is that the information from the contact form is no help whatsoever in deciding what the results of contact attempts by telephone are. From the information on the contact forms for the first wave it appears that in less than 5% of cases the interviewer noted down that they called up the respondent. In less than half of this they called only once. As it is not clear from the question of what the phone calls refer to, this information will not be used in the analysis.

The final shortcoming is that the reason for refusal was not keyed. One possible outcome of each call was that the sample household could not cooperate or did not want to cooperate, either because of a refusal or for another reason. In the event of a 'not able', the interviewer could put down when a new call would be possible. In the event of a 'not willing', the reason for refusal was to be recorded. This information was later used in deciding whether a reissue would be worthwhile (see previous section). Due to an unfortunate oversight, the contact forms were destroyed for reasons of confidentiality, before it became clear that the reason for refusal had not been keyed. What is available is the distinction between several transitory reasons for refusal (ill, engaged for different reasons) and one general reason for refusal, and the assessment of interviewers of the likelihood of success in a subsequent wave.

An additional problem with refusals is that on different occasions a different household member may have reacted differently to the request of the interviewer, or given different reasons for refusal. At the end of the day, however, every member of the household had to cooperate in order to provide a complete set of questionnaires. Thus if one member of the household refused and another agreed to cooperate later, the first one still had to fill in the paper-and-pencil questionnaire (or agree to have it filled in by the other household member). Keeter et al. (2000) encountered a similar problem in their analysis of the consequences of reducing nonresponse in a telephone survey. Accordingly, they refer to 'amenable households' and 'reluctant households' instead of amenable or reluctant respondents. Smith (1984, p. 485) ascribed his inability to explain temporary refusals partly to the fact that these temporary refusals might have originated from family members other than the selected respondent.

In the description of fieldwork and paradata collection in this section, particular attention has been paid to a number of shortcomings in the fieldwork and the collection of paradata. These inconsistencies and defects are negligible in analysing substantive surveys outcomes, but they hamper the computing of exact response rates and analysing the response process and nonresponse bias. In the subsequent chapters, however, it will be seen that, even though paradata could have been more complete, a wealth of information is available that allows nonresponse analysis in more depth than in most (Dutch) studies. In designing the fieldwork for the AVO2003, and also in the Dutch fieldwork for the European Social Survey 2002/2003, most of the shortcomings above have been mended. It will still be clear, however, that collecting standardized paradata is a cumbersome process.

That said, a number of recommendations can be drawn from inspecting the fieldwork and the fieldwork documentation of the AVO1999. These recommendations have been implemented in the avo2003. Here they will be presented in combination with paradata collection strategies that were already implemented satisfactorily in the AVO1999. In face-to-face surveys the following paradata should be collected:

- Independent information on the eligibility of sample addresses. The vacancy of unit should not be inferred from the failure of a possible resident to open the door. The contact form should distinguish different reasons for ineligibility, such as building demolished, building clearly uninhabited, second dwelling, etc.
- In an address sample, for multiple household units a random selection scheme should be developed similar to the selection procedure of an individual from a household. The number of households per unit and the number of households to be interviewed should be accounted for.
- When trying to contact a household all calls should be accounted for, either in person or by telephone, or contacts initiated by the respondents (when someone calls the survey agency after receiving an advance letter, to make an appointment or refuse to cooperate). Modes of calls should be distinguished.
- The time and date of each call should be recorded.
- Reasons for refusal should be recorded and keyed, and if possible who refused.
- The characteristics of the interviewer may play a role in the decision to cooperate. For each call an interviewer id should be known, and some interviewer characteristics: at least age and sex, possibly an indicator of experience and dependent on the type of survey an indicator of ethnic group.

6.3 Follow-up survey among refusers AVO1999

6.3.1 Aims and design of the follow-up survey

The primary goal of the follow-up survey among persistent refusers was to discover whether refusals differ and if so how , and to assess the impact of this difference on survey results. A secondary goal was to find out how very high response rates can be obtained in a somewhat unrealistic setting where time and money do not present

practical limits. To achieve these goals, a target cooperation rate of 80% among persistent refusers was set.[4]

According to the AVO1999 technical field report (GfK, 2002), 2521 householders had told the interviewer they would not (2339) or could not (182) participate. Of these, 1063 households had refused at least twice and 693 persons had made it abundantly clear that they did not want to be interviewed and had not been approached a second time. The total group of refusers thus amounted to 1756 households (765 nonrespondents remained who had refused at least once but had not been re-approached although the interviewer had not characterized them as absolutely unapproachable). These households were the target population of persistent refusals from which a sample of 350 households for the follow-up survey was drawn.[5] Respondents who had moved were not followed to their new address as the original sample was based on addresses, and characteristics of the dwelling and the neighbourhood would change after the move, unnecessarily complicating the analysis. Of the 350 addresses 21 were excluded because they could not be reached after a large number of calls (12) or were unoccupied dwellings (5), leaving a net sample of 329 refusers. They were contacted during the summer months of 2000.

To achieve high response rates among this difficult group, a rather single-minded strategy was used: pull out all the stops and implement all possible techniques suggested in the literature (section 4.3). Highly motivated interviewers with a history of high response rates were selected (and told why they had been selected) and trained, and received additional monetary incentives for each successful interview plus an additional bonus if they personally achieved an 80% response rate. Interviewers could spend an average of approximately EUR 22.50 on each respondent to buy flowers or gifts, pay for their time, and invite respondents to answer questions in a public place. They could spend the money as they deemed fit, but were asked to inform the fieldwork organization if a strategy was successful, so it could be recommended in a newsletter.

4 When evaluating this target it should be noticed that at least part of the persistent refusers would have cooperated if the design and burden of the original survey had been similar to the follow-up survey. On the other hand: obtaining answers to an AVO 'light' questionnaire from 80% of the refusers to a survey where the refusal rate was about 30% and about 65% of the households had given the requested information earlier, results in a response rate of almost 90% (65% + 0.8 x 30%) to the AVO 'light' question set, which is without precedence for a general face to face household survey in the Netherlands.

5 Actually, two random samples of 350 persistent refusers were drawn, the primary sample and a shadow sample. After inspection of the contact forms of the primary sample, it turned out that 15 households could not be classified as persistent refusers. They were substituted by 15 households from the same regions from the shadow sample.

Table 6.4 Overview of questions in the VO1999 follow-up survey among refusers (multi-mode) (p.: 5-pages)

subject	p.	subject	p.	subject	p.
family composition	2	cultural participation	4	contacts with police and legal aid	2
dwelling	1	media use	3	religion and ethnicity	2
health, personal care, domestic care	1	arts and crafts	½	membership of organizations	½
care for children	1	educational level	1	occupation and disablement	1
sports activities	2	use of counselling (labour market)	1	family income	1

Compared to the regular AVO the response burden of the follow-up survey was substantially lower; only one adult per household was approached. This is similar to the approach in the regular AVO, where one adult is asked to answer the face-to-face household questionnaire (see also Keeter et al., 2000). They received only a subset of the original AVO questions (see table 6.4) pertaining to background, use of services and issues presumed to be related to response behaviour (social and cultural involvement and participation). The household representative also had to answer a small number of questions concerning other household members. Respondents could answer face-to-face, by telephone or Internet or fill in a paper-and-pencil version. A wide range of rewards was available. As mentioned above, interviewers could and did give them tangible rewards, publications from the SCP and GfK, show them a newspaper article emphasizing the innovative character of the survey and, if they cooperated, they could if they wished be removed from the databases of survey and direct marketing organizations. To establish trust interviewers were able to use an introductory letter from SCP explaining the purpose of the survey and mentioning the earlier refusal. This letter had not been sent in advance in case it should backfire. They were also able to hand the respondent a card containing the direct telephone numbers of the directors of SCP and GfK, respectively. The combination of these measures led to a final cooperation rate among 'unshakeable' refusals, not of 80% but still of more than 70%.

To be able to compare the results of the 'light' AVO with the original AVO as regards the refusers, a control group was introduced to which the 'light' AVO questions were asked by the interviewers from the follow-up survey, according to the same design but without any special incentives. This independent, fresh sample of 276 addresses yielded 177 cases with which to spot questionnaire effects. The response rate among this control group was rather high (64%), considering the timing of the fieldwork (holiday period), the absence of special incentives and the limited number of calls (maximum four). The relatively high response rate may be due to the relatively low response burden (one informant per household, short questionnaire) and to the com-

mitment and dedication of the very motivated group of interviewers in the follow-up survey. In intervention research this type of 'leakage' would be considered a confounding factor; here the high response rate in fact facilitated controlling for design effects.

6.3.2 *Why did the follow-up survey succeed?*

Interviewers were encouraged to record their successful strategies as a source of information and inspiration for their fellow-interviewers, and were able to compare notes by means of a special newsletter. Their records give an impression of the incidence of different tactics. A first inspection of the records shows that interviewers were resourceful and frugal. They held interviews via the entry-phone, and visited respondents in hospital (after consulting the family). Several interviewers bought flowers, sometimes spending as little as EUR 2.50. A second finding of interest was that some respondents were very easily accessible and had no problem at all with being interviewed. From the contact forms it appears that in almost 70% of the successful approaches 'motivation' was used and that in slightly more than 70% of the cases monetary incentives were given, ranging from EUR 5 to 35, with the average being EUR 15, generally in the form of gift tokens. Sometimes money was paid into the account of the respondents, sometimes lottery tickets were given. In a small number of cases, for instance if the interview was administered by telephone, the interviewer called on the respondent with a gift voucher even if no money had been promised earlier. 70% of the interviews were conducted face-to-face, 20% by telephone, 8% paper-and-pencil and only 1% via the Internet questionnaire. 10% of the respondents asked to be removed from the databases of market research organizations so as to be free from future interviews and telemarketing.

Of those who refused to cooperate in the follow-up survey 37 (one third) agreed to be interviewed by telephone after a call back by the telephone unit from the organization. Reasons for non-cooperation were objections of principle and sometimes illness or the recent death of a partner. Remarks by interviewers on their (unsuccessful) efforts to persuade respondents to cooperate sometimes betray astonishment: *How can someone not respond in this survey when cooperation is so profitable?*

Three focus groups consisting of 24 of the 30 interviewers discussed the follow-up survey, interviewer tactics and nonrespondent characteristics. The interviewers had liked the follow-up survey, especially as they had almost reached the target. Several wondered why refusers had not cooperated in the regular survey as a considerable proportion of their allotment of hardcore refusers cooperated without any reservations at all. As major success factors they mentioned the higher than usual payment, the multi-mode character (choice between face-to-face, telephone, self-completion, Internet) and the possibility of giving (monetary) incentives.

Preparation and support

The initial briefing at which they had been confronted with the difficult task of converting 80% of the persistent refusers was very important. The fact that they had been identified and selected as the top 6% of the interviewer corps, that they had taken up the challenge, the trust placed in them and the resources they received (everybody left with a cash sum of EUR 22.50 per respondent to be spent as they deemed fit) greatly flattered their professional pride. The interviewers were in close contact with the survey organization. Their supervisor appeared to be well informed of the progress of each individual interviewer and their assignment of refusers. The support and counselling by GfK and the regular newsletter describing the status and presenting strategies that had been proved to be successful greatly helped and motivated the interviewers. They believed it could be done, except for one person who inspected the results of the former interviews, felt that the situation was hopeless and realized only four out of twelve interviews.

Doorstep behaviour

Several interviewers mentioned the significance of the first three seconds of interviewer-respondent interaction and the importance of keeping talking ('the longer the contact the likelier the success; the wider the door is opened the more easily you get in'). This is remarkably similar to an interviewer remark recorded by Sturgis and Campanelli (1998, p. 7): 'It's usually in those first few minutes, not even minutes, seconds. They open that door, they have almost said 'yes' or 'no' in their mind to the way you are stood at the door.' Almost all emphasized the value of thorough preparation: from the earlier waves they had information on the type of household, dwelling and neighbourhood, the age and sex of the refusing respondent and the reason for refusal. A number of interviewers adapted their appearance (clothing) to the neighbourhood they visited; others tried to look uniformly presentable. Interviewers who spoke the local dialect or were local residents felt this was an advantage. Almost all of them emphasized the detrimental influence of telemarketers on survey cooperation. The name and status of the sponsor of the survey (SCP) and the offer of the personal telephone number of the directors of both SCP and GfK added to the success of the survey, even though these telephone numbers were hardly ever used. Carton (1999, p. 13) emphasizes that the introduction of quality management in survey organizations will only be possible if the personal involvement of the top management is clearly visible. Several interviewers assured the respondents that great care would be taken to guarantee their privacy. Not many people wanted to be removed from sample frames and databases for future research and direct marketing. One respondent enjoyed the interview (and the reward) so much that he retracted his request to be removed from these databases.

Monetary incentives

The attitude of interviewers towards handing out substantial monetary incentives was ambiguous. On the one hand they appreciated the possibility of giving an incen-

tive and in some cases the money offered was the main reason for the respondent's willingness to cooperate. As might be expected, less affluent people and students responded positively to the offer of a cash incentive. On the other hand, the interviewers preferred respondents to participate because they felt the survey was important, and they felt cheap if they had to bargain over the amount of money to be paid; having to 'bribe' respondents wounded their professional pride. In addition, they preferred small cash incentives ('I felt 5 euros was more than enough for a 15-minute interview').

Means and efforts

Everyone agreed that they had tried harder than usual. As one respondent said: 'You will keep on calling unless I cooperate, won't you?' They did not feel that they had done anything out of the ordinary besides using the multiple modes of administration and incentives mentioned earlier. The availability of incentives had a positive influence in itself. Knowing that they could offer cash rewards boosted their self-confidence even if they did not actually proffer money. The same holds for respondents: in many cases the offer of an incentive was more important than actually receiving one. Several respondents in the end refused the incentive that had originally persuaded them to participate. In one or two cases the interviewees offered the gift voucher they had just received to the interviewer to buy themselves a present.

Pride and prejudice

The professional pride of the interviewers ('they will cooperate if I ask them') helped them to achieve a high response, but could also have prevented them from adopting alternative contact strategies. Despite the special instruction they had all received in advance, each of them seemed to have their own idiosyncratic tactics for obtaining cooperation and all seemed to succeed in their own way. Campanelli, Sturgis and Purdon (1997, p. 4-63) found in focus groups of experienced interviewers that successful strategies often appeared to be very specific to individual interviewers. They concluded (in bold): 'The key thing here then is that rather than going for a particular proscribed approach, interviewers should instead concentrate on honing an approach which is compatible with them as individuals and with which they feel comfortable, their own 'tricks of the trade.'

The interviewers in the follow-up survey did not consider their allotted households as adamant refusers. They ended up by redefining hardcore refusers as those whom they themselves could not persuade to cooperate. Hardcore refusers were not the group allotted to them, but the final group of refusers (30-34%) who did not want to cooperate in the follow-up survey. In these cases even an incentive of EUR 400 would not suffice. Surprisingly, as mentioned earlier, no fewer than a third of these 'hard' refusers could be induced to take part in an interview by telephone when they were called by the telephone unit from the survey organization. The dedication and trustworthiness of the interviewers made this difficult for them to accept. In a

number of cases where they had failed and the telephone unit succeeded, they called again on the respondent to hand over an incentive and ask why the respondent had not cooperated with their request. Finally, female interviewers expected men to have more problems when interviewing (elderly) women. The selected group of successful interviewers, however, contained a slight overrepresentation of male interviewers.

Difficult nonrespondents

The interviewers found the following types of respondents difficult to persuade, for different reasons:

- elderly women living alone and afraid to let someone inside (though some (male) interviewers had no problems at all with elderly ladies and were generally welcomed warmly);
- double-earner families with small children (too busy, no time);
- the seriously rich, for whom no incentive was of any value and who were too busy earning money;
- lower-class people who did not want to be bothered, were suspicious of governmental interference and distrusted the interviewers;
- inhabitants of apartment buildings with entry-phone (difficult to get inside, interviewers brushed off through the entry phone).

Highly educated or upper-class people presented a different kind of problem, again according to the interviewers. They either cooperated or they did not cooperate. If they did not, eloquence and incentives were of no avail: they refused in a friendly but determined manner and did not feel obliged to explain why, which vexed the interviewers considerably. Family interaction could also be a problem: in several cases one of the spouses would have cooperated if the other had not prevented this. One interviewer recorded as a reason for a broken-off interview: 'Stopped during the interview under the influence of the spouse: 'See, they want to know about our possessions and we are being harassed again'. Scared of burglars'. Another one put down: 'Woman in doubt, man refuses.' A third one decided to refrain from reapproaching the sample household, because the husband had just died ('although he was the one who refused in the original survey').

Success factors

How was it possible that almost three-quarters of adamant refusers were willing to cooperate in the follow-up survey? From the respondents' point of view the lower response burden certainly helped, as did the option of choosing another interview mode. For at least some of them the sizeable financial incentive may have made a difference, or the promise that they would not be bothered again. The focus group results indicate that the respondents were very well aware that the interviewer wanted their cooperation badly. From a theoretical point of view the distinction made by Groves and Couper (1998, p. 38-39) between two different optimization targets in developing an introductory strategy may illustrate the difference between

interviewer efforts in the follow-up survey and in ordinary household surveys. The predominant optimization target of the sampling procedure of Dutch surveys comes close to quota sampling interviewing: 'An acceptance of the survey request is preferred to a denial, but a lengthy, multicontact preliminary to an acceptance can be as damaging to productivity as a denial.' In the follow-up survey the second optimization target described by Groves and Couper was used, viz. probability sample interviewing in which the probability of each sample unit accepting is optimized: 'The odds of success are increased with the continuation of the conversation. Thus the interviewer does not maximize the likelihood of obtaining a 'yes' answer in any given contact, but minimizes the likelihood of a 'no' answer over repeated turn-taking in the contact.'

Thus the persistence of the interviewers may have helped. They were highly motivated and no means were spared to give them a sense of self-efficacy, as they were handed money and a wide range of non-monetary incentives. In addition, they were told they were top interviewers and extensively briefed on the aims and the importance of the follow-up survey (see also section 3.4.3). It seems likely that the combination of high-quality, highly motivated interviewers, a wide range of incentives from which to choose and information on the respondents and earlier interactions led to the success of the follow-up survey. Interesting is the finding mentioned earlier that the interviewers did not feel happy about giving interviewees substantial amounts of money. In their opinion, a monetary gift should be a token of appreciation and so they preferred to hand over a gift of EUR 5, which corresponds quite nicely with the USD 5 that Singer et al. (1999) found effective.

After hearing the discussions in the focus group, it seemed likely that a large part of the effect of the incentives was mediated by the interviewers: the wide range of tangible and intangible means they had at their disposal might make participating more attractive for respondents, but was also a way of showing interviewers that their work was really important and worthwhile, and of increasing their motivation (see also section 3.5.6). They were very happy about the possibility of handing out incentives, and felt that this may have helped in achieving high response rates.

When comparing the procedures and the evaluation of the follow-up survey to the social exchange framework described in section 4.3, one might conclude that it is not possible to identify the single most effective strategy from the follow-up survey and adopt this in other surveys. What will have been a major factor, and one which permeates all details of the fieldwork, is the total dedication and motivation of the sponsor, survey organization, survey staff and interviewers to make this work. Another key factor will have been that, while costs were kept to a minimum, for every refusal there might have been a specific benefit, such as a substantial monetary incentive, the knowledge of being part of an important scientific endeavour, helping out an interviewer who went to great lengths to obtain cooperation, etc. Section 4.4

argues that social exchange is not so much a theory from which hypotheses can be derived, but more a framework for obtaining survey cooperation. Implementing the evidence from this framework turned out to be very effective. Successful fieldwork strategies from the follow-up survey have already been implemented in other surveys, e.g. the Dutch first round of the European Social Survey.

Differences between original survey and follow-up survey

The follow-up survey was set up to measure bias due to nonresponse in the original survey. This required a high response rate. To achieve this high response rate the response burden was reduced considerably. It could be argued that the cooperation rate in the original survey would have been much higher if respondents had been asked to cooperate in the 'light' AVO from the beginning. Of course, the contents of the survey would in that case have been much less relevant (see also Fellegi, 2001, for a discussion of the trade-off between survey quality criteria). As mentioned above, the shorter questionnaire (15 minutes instead of half an hour) and the fact that there were no drop-off questionnaires for each member of the family certainly helped. However, even in an individual survey compared to a household survey, cooperation will sometimes be a family decision.

Another factor that may partly explain why refusers decided to participate in the follow-up survey is that different modes were possible. According to the interviewers, the multi-mode design was one of the success factors of the follow-up survey. More than two-thirds of the interviews were conducted face-to-face (as in the original survey) and almost a quarter by telephone (two-thirds by the central telephone unit, not by the interviewer). The Internet version was selected by less than 1%; the remaining respondents filled in the questionnaire themselves. Because of the very small numbers and the factual nature of the questions, mode effects have not been studied.

Differences between respondents in the follow-up and original survey could be due to survey and questionnaire design differences. The control group that answered the questionnaire concomitantly to the follow-up survey made it possible to isolate these design effects. It turned out that a number of questions could not be used in comparing respondents and nonrespondents, because the control group showed that abridging and slightly rephrasing questions had a large impact on the results. These questions have been excluded from further analysis (see chapter 9).

6.4 *Data file empirical analysis*

The AVO is a household survey comprising information on the household, provided face-to-face by a knowledgeable adult, and on all household members, provided by all members of the household through a paper-and-pencil questionnaire. In the follow-up survey only one person per household participated, namely any knowl-

edgeable, cooperative adult who could provide information on the household and on him or herself, either face-to-face or by telephone, or self-administered (paper-and pencil or Internet). To compare these surveys, the data files were harmonized by selecting from the regular AVO only those respondents who answered the household questionnaire. The household data were added to their individual data. As a consequence, the comparison is between the respondents from the follow-up survey and those regular AVO respondents who would have been in the file if only knowledgeable adult person per household had been interviewed. The data may contain an overrepresentation of those household members who are the most housebound and, as a consequence, will not represent the entire Dutch population. This is for comparison purposes only, as in the regular AVO everybody had to cooperate.

It should furthermore be noted that all analyses were performed on the unweighted data. The unit of analysis is 'household', and as it is an address sample design weights would only have been necessary to correct for the very small percentage of multi-household addresses. As this is not completely documented (see chapter 5), no design weights are used. Nonresponse weights would of course defeat the purpose of the analysis. Variables that are incorporated in the nonresponse weights of the final substantive analysis file, are generally included in the model as independent variables.

The analysis file that will be used in the subsequent chapters comprises the substantive data shared by the regular survey and the follow-up survey, data from the geographical databases discussed in section 6.2, and a regional indicator (see figure 6.1). In some cases only the largest cities (Amsterdam, Rotterdam and The Hague) are compared with the less urban areas. Figure 6.1 gives a geographical overview, distinguishing between the most highly urbanized areas (the three largest cities Amsterdam, Rotterdam and The Hague in the western part of the Netherlands), the urban remainder of the western region, and the somewhat more rural other regions.

Figure 6.1 Regions in the Netherlands

● Am erdam, Rotterdam, The Hague
West, other: Noord-Holland, Zuid-Holland, Utrecht (excl. big cities)
North-east: Groningen, Friesland, Drente
Mid-east: Gelderland, Overijssel, Flevoland
South: Zeeland, Noord-Brabant, Limburg

7 Ease of contact

7.1 Introduction

Ease of contact in face-to-face surveys is a function of the call pattern of the interviewer, the presence of impediments, and the size and at-home pattern of the sample household, as outlined in the model of contactability in figure 2.3. The process of establishing contact with sample units was described in section 2.6.2. Section 3.2 identified those groups in society that are generally considered to be hard to contact and in section 5.2 a continuum of resistance was presented as a scale on which to position increasingly hard-to-reach respondents. The present chapter will put into practice the findings of these earlier studies.

As argued in section 2.6.4 the best – but still imperfect – indicator of ease of contact is the number of calls[1] to first contact. The reason why it is imperfect will be illustrated in section 7.2. The indicator requires detailed call records that are generally not available but were kept and keyed in the AVO1999 survey (see section 6.2.3). Section 7.2 will also discuss the timing of calls in the AVO, the contact rate of calls at specific times of the day and on specific days of the week, and the cooperation rate of calls at which first contact was established. Strictly speaking, the latter issue falls outside the scope of a chapter on contactability; on the other hand, the question of whether the timing of calls influences contact is very closely related to the question of whether contacts at particular times may induce less or more reluctance. Finally, section 7.2 will introduce telephone calls as an additional means to establish contact, and the relationship between the availability of a listed phone number and socio-structural characteristics of households. Section 7.3 will investigate the existence of a continuum of resistance and identify those individuals and households that are hard to reach. The analyses of contactability will be similar to those of Campanelli, Sturgis and Purdon (1997), Groves (1989), Groves and Couper (1998) and Lynn et al. (2002). Section 7.4 will span a bridge between this chapter on ease of contact and chapter 8 on reluctance to cooperate and again step outside the limits of contactability as such. It will compare total field efforts and efforts to contact sample units and analyse whether contactability and cooperation are independent factors in survey participation. Section 7.5 will discuss the outcomes of this study with respect to contactability and give suggestions for improved contact forms and efficient fieldwork procedures.

1 Throughout this study a call is a personal visit to a sample address. In this particular survey telephone calls were allowed in later stages of the fieldwork. These will be identified as such, if possible (see section 6.2.2).

7.2 Calling on the sample households

7.2.1 Timing of calls and contact rates

In preparing fieldwork, interviewers are generally instructed to spread their calls over the hours of the day and the days of the week to counteract standard at-home patterns of the respondents. If someone is not at home on a Monday morning, she might have a paid job and be not at home at other mornings too. And if someone is not at home on a Monday evening, he might have his regular ballroom dancing class and therefore never be at home on Monday evenings. Spreading fieldwork over a longer period will also be useful in getting in touch with persons who spend prolonged spells abroad or in the hospital (see section 2.6.2). To monitor the spread of fieldwork, detailed paradata are required that record the precise timing of each individual call. In chapter 6.2.3 it was stated that the call records of the AVO1999 only contain month, time of the day and day of the week, but not the exact date. As a consequence, the effect of the time lag between calls could not be analysed, nor was it possible to ascertain when a subsequent call ought to be made dependent on the timing of the previous call. What can be done is to analyse the timing and contact rates of subsequent calls.

The timing of calls is presented in the upper part of table 7.1. This table contains only calls to households that have not yet been contacted at previous calls, as otherwise the timing of the calls could have been arranged with the households. Clues from earlier unsuccessful calls (comments from neighbours, evidence of the presence of small children) may of course have affected the timing of subsequent calls (Groves and Couper, 1989, p. 82) which is one of the reasons why the number of calls to first contact is an imperfect measure. Later calls (6 to 12) to noncontacted households are not shown owing to the small number, as only about 400 households needed more than five calls. Afternoons are rather popular, especially at the first call: more than 60% of the first calls to a household take place in the afternoon, 25% in the morning and 14% in the evening. The number of evening calls at previously non-contacted households increases from 14% at the first call to 27% at the second, 36% at the third and 38% at the fourth. After the fourth call the percentage of evening calls decreases. The spread of calls over days of the weeks shows no clear pattern except for an increase of calls on Saturday.

The bottom section of table 7.1 presents the success rate of calls at households that have not previously been contacted. If the first call is in the morning, the contact rate is 37%, in the afternoon 48% and in the evening 56%. At subsequent calls the same pattern holds: in the morning the contact rate is lowest, in the evening highest. Differences in contact rate between days of the week do not show a clear pattern. At the first call, for instance, Fridays have the highest contact rate (50%) but at the third call the lowest (40%).

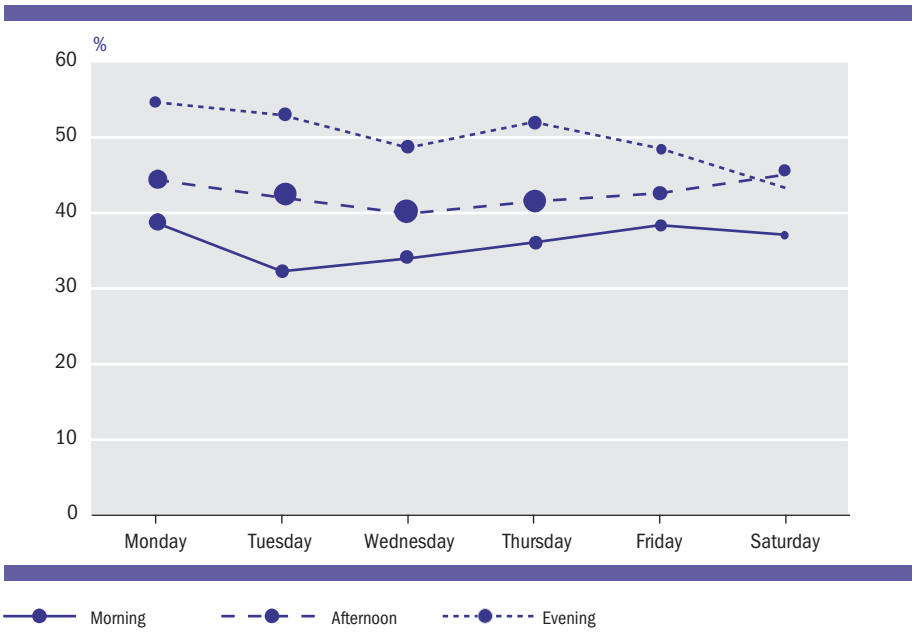
Table .1 Timing of first 5 calls and all calls to previously non-contacted household and contact rate (%)[a]

	1	2	3	4	5	all calls (1 to 12)
timing calls						
morning	25	27	26	24	23	25
afternoon	61	47	38	38	54	52
evening	14	27	36	38	23	23
total (N=100%)	9037	4744	2684	1380	763	20083
Monday	22	18	20	18	15	20
Tuesday	20	20	17	20	22	20
Wednesday	21	21	20	18	22	21
Thursday	19	20	18	18	16	19
Friday	13	13	16	16	14	14
Saturday	6	8	8	10	11	7
total (N=100%)	9157	4840	2768	1438	773	20467
contact rate						
morning	37	35	42	40	26	36
afternoon	48	38	41	37	28	42
evening	56	52	52	57	47	51
total (incl. period missing)	48	42	47	47	33	44
Monday	48	49	46	48	36	46
Tuesday	45	39	48	53	34	43
Wednesday	46	38	43	37	24	41
Thursday	46	41	49	47	36	43
Friday	50	41	40	51	25	43
Saturday	47	43	48	44	40	43
total (incl. day missing)	47	42	45	47	32	43

a Results of first five calls, 'all calls' presents the aggregate results of call 1 to 12

Figure 7.1 presents the contact rate of calls to previously noncontacted households dependent on both time of day and day of the week. The larger the bullet, the more calls were made within a particular time slot (many on Thursday afternoons, very few on Saturday evenings). The figure again illustrates that afternoons are very popular (large bullets) and that the contact rate is highest for evening calls, except on Saturdays. The number of Saturday evening calls is very small, however (0.63% of the twenty thousand calls made to first contact). As the bulk of the first calls are made in the afternoon, this implies that – despite the lower contact rate than during evening hours – many contacts are established during the afternoon. These results suggest that if someone wants to avoid being interviewed, they should make sure they are not at home on Monday, Tuesday, Wednesday and Thursday afternoons, because that seems to be the period when interviewers are most active. That will not ensure a permanent reprieve from being interviewed, however, as during fieldwork interviewers

Figure 7.1 Contact rate across times of the day and days of the week



shift their call patterns to evenings, and also to Fridays and Saturdays. The pattern of calls and contact rates across times of the day are similar to those found by Campanelli, Sturgis and Purdon (1997, p. 3-9) although the present contact rates are lower and the differences between mornings and afternoons much greater.

The possible presence of unrecorded telephone calls during which appointments were made in the AVO1999 complicates the analysis of contactability. According to table 7.1 the percentage of evening calls reaches a maximum at the fourth call. This may be a consequence of the contact strategy described in section 6.2.2. If the fourth call occurred in the first fieldwork wave, it may have been preceded by telephone contact between the interviewer and the respondent that could not be recorded on the contact form. As a consequence, a call at this stage may have been in the evening because the respondent preferred to be interviewed at that time. Table 7.2 suggests that this may well have been the case. It presents those previously noncontacted households that received a fourth call in the first wave (and thus possibly after a phone call). The percentage of evening calls is very high here, as is the contact rate. This suggests that a part of the evening calls in table 7.1 may have been preceded by appointments made by telephone.

An additional confounding factor may be the possible relationship mentioned in section 2.6.2 between local knowledge and specific fieldwork strategies of interviewers, and characteristics of respondents. In the present survey interviewers tended to call upon sample households that lived in inner cities, neighbourhoods with mainly

Table .2 Timing calls and contact rates for fourth call, first wave[a]

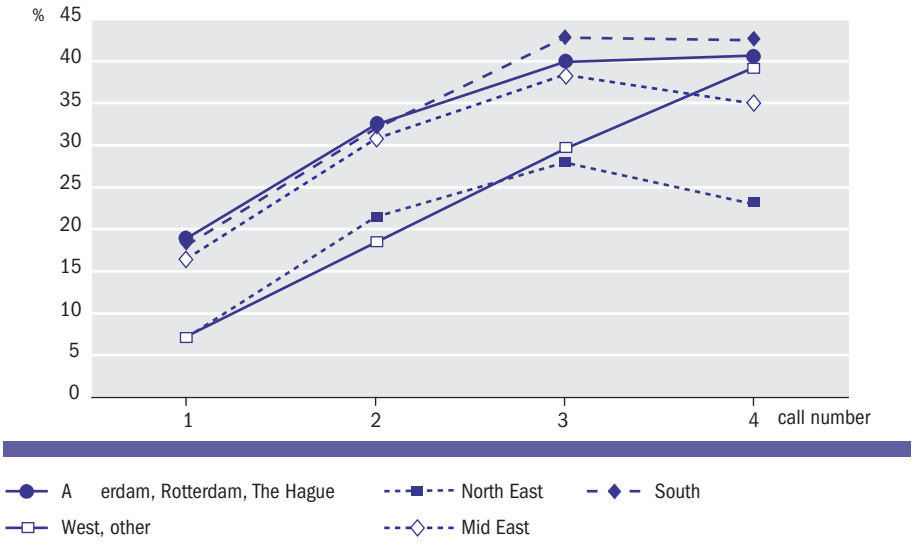
	% calls	contact rate
morning	20	50
Afternoon	28	54
evening	52	64
total (=100%)	732	58

a Only those calls to previously noncontacted household of which the time was known

high-rise buildings or multi-unit structures more often in the evening and less often in the morning. The differences were small, however, and not at all clear-cut. For instance, call patterns in the western, urbanized region of the Netherlands differed between the three major cities (Amsterdam, Rotterdam and the Hague) and the remainder of the region (see figure 7.2). Interviewers in these big cities made slightly more evening calls than in other regions, and interviewers in the remainder of the region showed a similar pattern to that of interviewers in the rural north-eastern part of the Netherlands, at least during the first three calls.

It will be clear from this section that the contact rate differs depending on the time of day and that the number of calls to first contact can be reduced by making more evening calls. This may not reduce the length of fieldwork, however, because only a certain number of evening hours are available; nor may it reduce fieldwork costs, as evening calls may be more expensive and it will be more difficult to combine calls on different sample persons. In addition, it the possibility cannot be ruled out that interviewers have adapted their calling pattern to local circumstances, which may be beneficial for the contact rates but complicates the analysis of contactability.

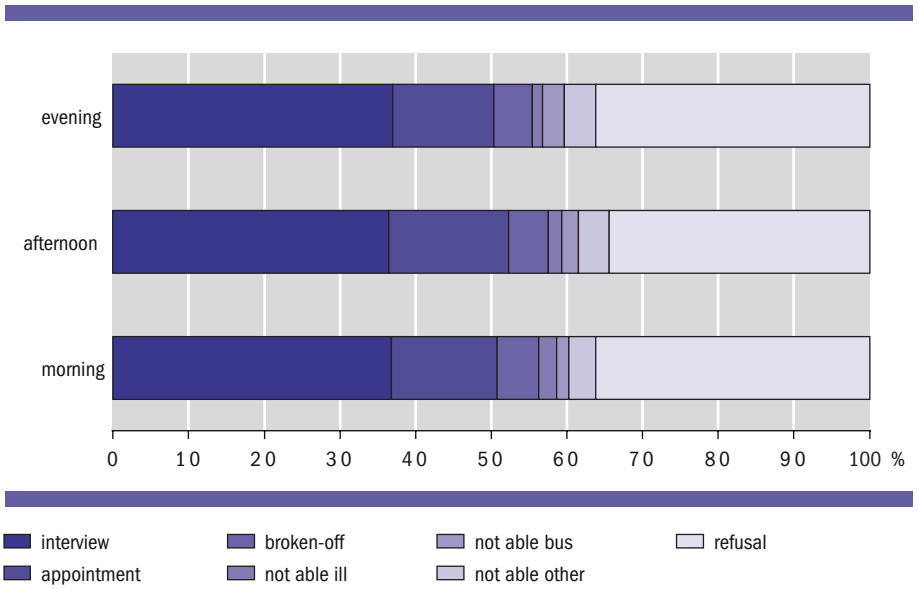
Figure 7.2 Percentage of evening calls to previously non-contacted addresses per region



7.2.2 *Timing of calls and success of contacts*

The previous section has shown that the best time to contact a household is during the evening. This would be a self-defeating strategy if these evening contacts more often resulted in a refusal. A similar dilemma is often cited with regard to making phone calls to establish contact with households, as this might make contact easier but also make refusing easier for the contacted person (see below). Figure 7.3 presents information on the relationship between the timing of calls and the willingness to cooperate and distinguishes between outcomes of calls at different times of the day and different days of the week. Figure 7.3a shows the call outcome for all calls to first contact. The percentage of interviews is very similar, the percentage of appointments slightly higher for afternoon calls, and the percentage of sample persons not being able to cooperate because they were ill slightly lower for evening calls. This is probably not due to poor labour morale in the Netherlands, but to the presence of other healthy household members during evening hours.

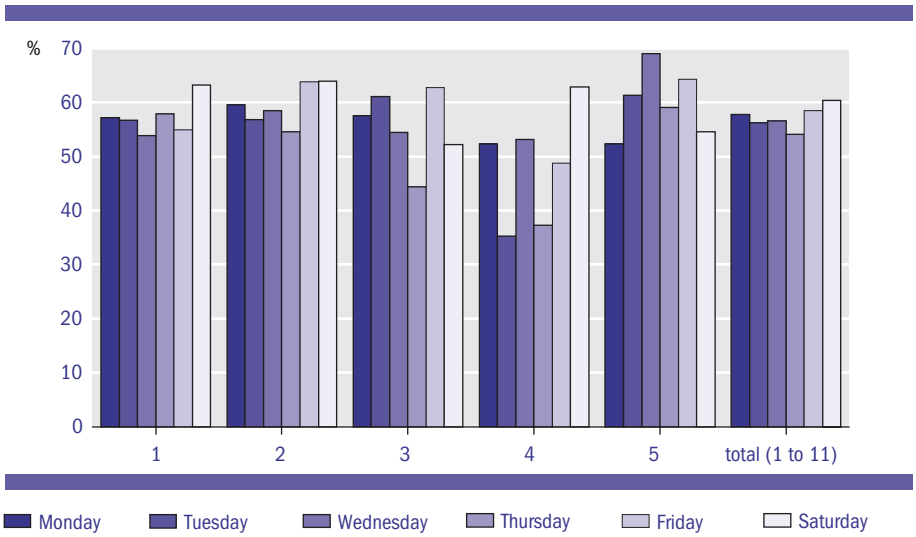
Fig re 7.3a Outcome calls per time of the day (all calls to first contact)



Figures 7.3b and 7.3c portray the likelihood of success for contacts that have been made during the first five calls and, in general, by time of the day and day of the week. If contact was made at the first call, the cooperation during the evenings was higher than during the daytime, if contact was made at the second call it was lower. Why these differences occur is difficult to interpret. The differences in cooperation rates at different times of the day at the fourth call may again be due to previous, unrecorded appointments. The results for all calls are identical to those in figure 7.3a and show that differences between cooperation rates for first contacts that have been made at different times of the day are very small. These results indicate that the

higher contact rate during the evenings is not counterbalanced by a lower cooperation rate.

Figure 7.3c Percentage initial cooperation (interview, appointment, broken off) per call to first contact and day of the week

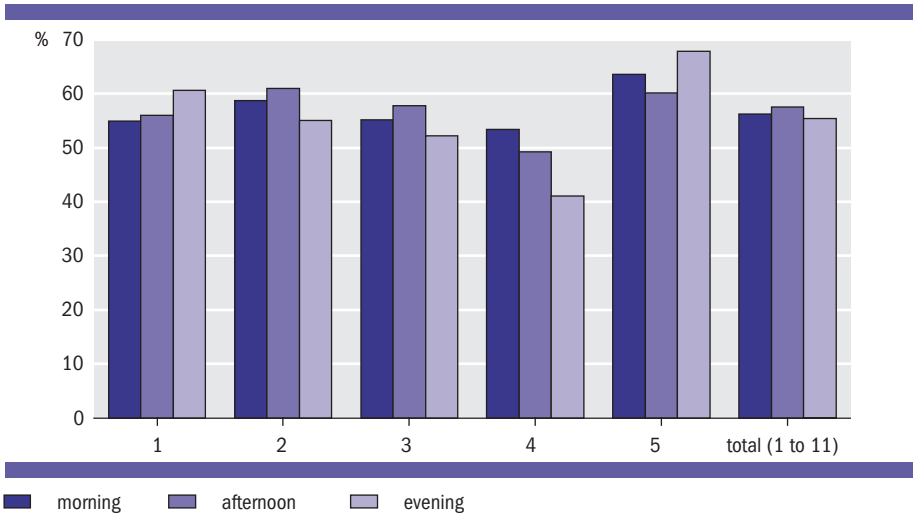


The differences between days of the week per call to first contact in figure 7.3c are much larger than for times of the day. These differences, however, are hardly consistent. Probably the only conclusion that can be drawn from figure 7.3c is that, if first contact is made on a Saturday, sample households are by no means less likely to cooperate than on weekdays.

7.2.3 *Telephone recruitment*

It is generally assumed that respondents are easier to reach by telephone but also refuse more often (see section 2.5). In the AVO telephone calls could be made to previously noncontacted households in later stages of the fieldwork (section 6.2.2). The availability of a listed phone number may have facilitated contact with hard-to-reach households. As the presence of a listed phone number may be unevenly distributed across the social strata of Dutch society (section 3.2) this procedure may have caused bias, although those addresses without listed telephone numbers were reapproached face-to-face in later waves of the fieldwork. Therefore, before incorporating the possibility of contacting a household by telephone as a determinant of contactability in the next section, the present section will analyse the relationship between 'telephone' and socio-structural variables of both sample households and respondents as measured in the AVO1999 (see Stoop, 2002b).

Figure 7.3b Percentage initial cooperation (interview, appointment, broken off) per call to first contact and time of the day



In the logistic regressions (see annex 7.1 on how to interpret the results) presented in table 7.3 the characteristic 'listed telephone number available' is predicted for eligible sample households by general information on the neighbourhood and the dwelling, and for respondents by environmental characteristics, background variables and survey variables. The analysis of the eligible sample households in the upper part of the table shows that the availability of a listed phone number is related to the maintenance of both dwelling and neighbourhood and to urbanicity. The negative and significant β values indicate that listed telephone numbers are less often available when maintenance is poor, when the proportion of ethnic minorities in the neighbourhood is high (see also section 3.3.4), when the dwelling is in one of the three major Dutch cities and when the neighbourhood is characterized by high-rise buildings. The exp (β) values represent the odds-ratios of each category having a listed telephone number compared with the other households. The lower section of the table presents the results of a similar logistic regression for the AVO respondents. Maintenance and urbanicity (including the percentage of ethnic minorities) are significant here, too, in addition to household and individual characteristics related to social and cultural participation, media use and sports activities. Among respondents, living in a detached dwelling indicated a higher probability of having a listed phone number but the presence of high-rise buildings in the neighbourhood was not significant. A large effect was found for respondents belonging to one of the four major Dutch ethnic minority groups. They comprise about 2.5% of the total AVO net sample, 6% of the households living in neighbourhoods with 16-25% minorities and almost 9% of those in neighbourhoods comprising more than 25% ethnic minorities, both according to the frame data. It is not only neighbourhoods with a high proportion of ethnic minorities that are characterized by a low proportion of listed phone numbers, therefore, but also these ethnic minorities themselves. Elderly people and the highly educated more often have a listed phone number, singles less often.

Table .3 Logistic regression availability listed phone number (upper part: eligible households, lower part: respondents)[a]

eligibles	%	%phone	β	exp β	standard error	Wald	sign. (df 1)
Amsterdam, Rotterdam, The Hague	16	65	–0.194	0.823	0.075	6.64	0.010
poor maintenance	21	63	–0.658	0.518	0.057	131.23	0.000
ethnic minorities 16+ %	18	62	–0.457	0.633	0.070	42.32	0.000
high rise buildings	35	69	–0.269	0.764	0.058	21.81	0.000
detached dwelling	14	90	0.788	2.200	0.098	64.95	0.000
constant			1.523	4.584	0.038	1578.94	0.000
total (n 100%)	n 9409	77					
model	X^2 (df 5)	529		Cox & Snell R^2		0.055	
–2 Log likelihood	9648			Nagelkerke R^2		0.083	
respondents							
Amsterdam, Rotterdam, The Hague	16	66	–0.433	0.649	0.094	21.00	0.000
poor maintenance	20	65	–0.584	0.557	0.077	58.19	0.000
ethnic minorities 16+ %	18	65	–0.351	0.704	0.094	14.11	0.000
detached dwelling	14	93	0.918	2.504	0.141	42.30	0.000
ethnic minority group[a]	3	34	–1.331	0.264	0.185	51.53	0.000
family composition (df 2)						18.06	0.000
single	27	74	–0.160	0.853	0.089	3.21	0.073
not single, no children	34	84	0.213	1.237	0.085	6.22	0.013
not single, children	38	77					
higher education	18	83	0.417	1.518	0.097	18.50	0.000
sports activities (1 or more)	60	80	0.208	1.232	0.074	7.87	0.005
household owns pc	56	80	0.196	1.216	0.080	6.07	0.014
watches sport on tv	63	80	0.131	1.139	0.068	3.63	0.057
	mean	mean phone					
age	49.3	50.7	0.026	1.026	0.002	113.88	0.000
membership organization	1.42	1.48	0.044	1.045	0.024	3.40	0.065
constant			–0.107	0.899	0.155	0.47	0.491
total (n 100%)	n 6125	79					
model	X^2 (df 13)	634		Cox & Snell R^2	0.098		
–2 Log likelihood	5695			Nagelkerke R^2	0.153		

a Country of birth respondent, respondent's mother or respondent's father Morocco, Turkey, Surinam, Dutch Antilles or Aruba.

The more organizations someone is a member of, the higher the likelihood of them having a listed phone number. The same holds for sports activities, watching sports on TV and owning a PC. These results point to a relationship between social involvement

and listed phone numbers and also seem to confirm the middle-class image of society that telephone surveys provide as a result of undercoverage and nonresponse bias (Van Goor and Rispens, 2004).

In the next section the absence of a listed number will be included as a possible impediment to contacting households. One complication is that telephone calls were only allowed in later stages for hard-to-contact households. As a consequence, the fact that a household has been contacted by telephone implies low contactability, whereas the fact that a household can be contacted by telephone may enhance contactability (see section 5.2).

7.3 *Continuum of resistance: characteristics of hard-to-reach respondents*

By now it is known when people were most likely to be found at home (figure 3.1), when interviewers called upon the sample households (section 7.2.1), which time slots had the best contact rates (table 7.2) and who could be reached by telephone (section 7.2.3). In addition, figure 7.4 depicts the efforts it took to contact eligible sample households in general. This section will investigate whether there is a continuum of resistance with respect to contactability. The call pattern, background variables and survey variables that may be related to contactability will be taken into account. Section 7.3.1 studies the contactability of all eligible contacted sample households and section 7.3.2 of the respondents. This group may differ from the total sample (see the next chapter) but is especially interesting as more information is available on respondents than on nonrespondents.

Figure 7.4 Number of calls to first contact, eligible households

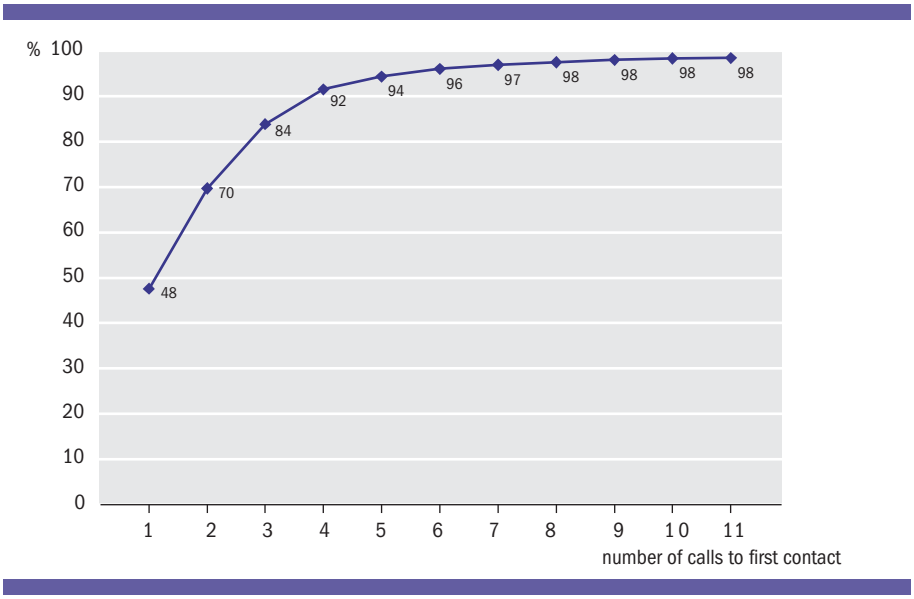
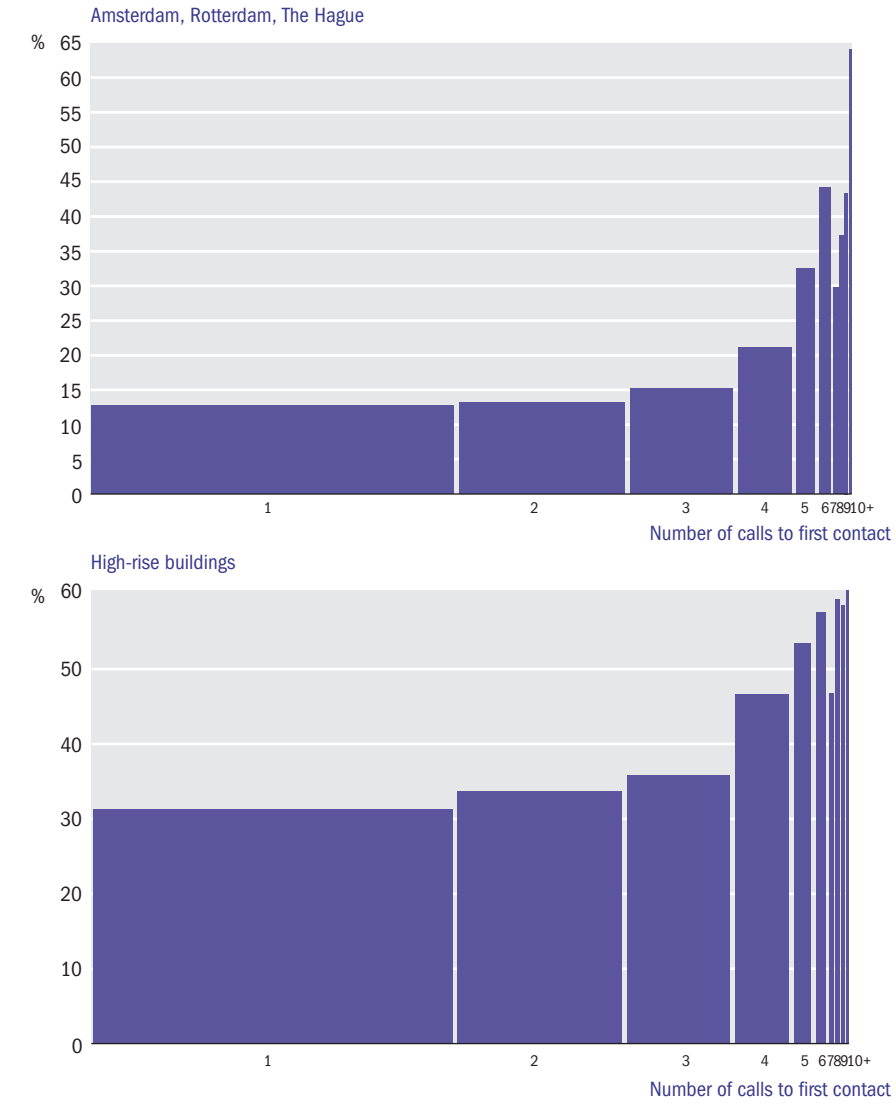


Figure 7.5 Number of calls to first contact and percentage big city (above) and high-rise buildings in neighbourhood (below), eligible sample units (including language problems)



7.3.1 Contactability of eligible households

In section 3.2 urbanicity was identified as one of the major correlates of contactability. Figure 7.5 shows how this worked out in the AVO.[2] This figure is a 'spineplot', a

2 Persons with language problems are incorporated in this figure, as they had to be contacted in order to be able to decide that they did not have to be interviewed.

variation of a bar chart where the width of the bar represents the proportion of cases in a specific group. The upper graph depicts the percentage of sample households that live in one of the three major cities in the Netherlands (Amsterdam, Rotterdam and The Hague) at each call to previously noncontacted households. It shows that the percentage of big-city dwellers is around 12.5 % at the first two calls, and more than 60% for those households that required 10 or more calls to be contacted. As the number of households that required so many calls is very small compared to the number of households that were contacted at the first or second call, the bar for 10+ is very narrow and that for call 1 and 2 very broad, in proportion to the number of households that were first contacted at each subsequent call. These results show clearly that the more calls required, the higher the urbanicity, although the pattern becomes quite fuzzy for those households that needed more than six calls to be contacted. Put differently, if fieldwork would have been cut off after two or three calls, big-city dwellers would clearly have been underrepresented, assuming that the call pattern in such a restricted design would have been similar to the current one (see the reservations of Curtin, Presser and Singer, 2000, presented in section 5.2). The lower graph in figure 7.5 presents in a similar way the percentage of sample households residing in a neighbourhood with high-rise buildings. The percentage of high-rise buildings in the neighbourhood increases with the number of calls required to establish a first contact up to the sixth call, then drops sharply at the seventh call before rising again for calls 8, 9 and 10.

Table 7.4 presents a more systematic approach and gives basic statistics and the results of logistic regressions for the first five calls to previously noncontacted households with complete contact forms. The first row shows that initially 9037 households were called on, 14% of them during evening hours and 6% on Saturdays. Of these households 23% had no listed phone number, 16% lived in a big city, 35% in an area with mainly high-rise buildings, 14% in a detached dwelling and 21% in a poorly maintained dwelling and/or neighbourhood. The second row shows that a second visit was made to 4683 addresses, which means that 9037-4683 = 4354 had been contacted at the first call. The percentage of households that were called on in the evening is 14% at the first call, 27% at the second, 36% at the third, 38% at the fourth and 23% at the fifth[3]. The sample composition at the second call to previously noncontacted households differs from the composition at the first call, because the contact rate is related to urbanicity. The upper five rows of the table show that, at each subsequent call, the remaining noncontacted part of the sample is increasingly an urban sample: the percentage of inhabitants of Amsterdam, Rotterdam and The

3 These results are identical to those presented in table 7.2.

Hague increases from 16% at the first call to 36% at the fifth call[4], the addresses in neighbourhoods characterized by high rise buildings from 35% to 55%, etc. The percentage of households without a listed telephone number clearly reflects the lower contactability of those who remain noncontacted after several calls.

The second five rows of the table represent the contact rate for the groups identified in the table header at each successive call. At the first call the average contact rate is 47%. This decreases to 41% at the second call, rises at the third and fourth call (45%, possibly due to telephone contact) and decreases to 31% at the fifth call. On each occasion, the contact rate of evening calls is substantially higher. Saturday calls, on the other hand, do not make a large difference. The contact rate in big cities and high-rise neighbourhoods is always lower than the average contact rate, which is of course why the remaining sample comprises more big-city dwellers after each subsequent call. The contact rate for households with no listed phone number is also lower than the general contact rate per call.

The remainder of this table presents the results of five separate logistic regressions. The first row under β represents the β values of a logistic regression of the dependent variable 'contact (yes)' at the first call to previously noncontacted addresses, and the first row under exp (β) the values of the odds-ratio. The first row under Wald gives the value of the Wald statistic, the first row under significance the significance of the Wald statistic, the first row under standard error the standard errors and the first row under χ^2 the difference between the final and the reduced model (see annex 7.1). The next rows in each block present logistic regressions of increasingly smaller and more difficult to contact groups up to the 751 households in the fifth row of each block that were visited for a fifth time and had not been contacted previously. The analyses, and those in table 7.5, are similar to those presented by Groves and Couper (1998, section 4.6).

The results of the logistic regressions are fairly unequivocal. Evening calls have a substantial positive effect on contactability for all calls. As could be expected, considering the contact rates mentioned above, the effect of Saturday calls is small, or nonexistent. The effect of the absence of a listed telephone number is especially large at the fourth call, when the interviewer could telephone the as yet noncontacted sample households. This suggests that the absence of a listed phone number is more of a practical impediment to contacting a household than a general indicator of being hard to reach. Indicators of urbanicity (being a resident of a big city, high rise-buildings in the neighbourhood, no detached dwelling and poor maintenance) have an adverse effect on contactability, particularly for the first three calls.

4 The outcomes in table 7.4 differ from those depicted in figure 7.5. This is partly because in the logistic analyses persons with language problems were excluded, as they had been excluded from the target population. More importantly, in figure 7.4 urbanicity is computed for those who were contacted at the first call, the second call, the third call, etc. In table 7.4 urbanicity is given for all previously noncontacted households that received a first call, a second call, a third call, etc. In figure 7.5 the results of the different calls add up to the eligible households; in table 7.4 all households receive a first call and the number of households that receives subsequent calls gradually dwindles.

Table .4 Logistic regression contact rate per call on households not previously contacted; eligible households[a]

number call	total/ constant	evening call	Saturday call	no listed phone number	msterdam, Rotterdam The ague	highrise buildings	detached dwelling	poor mainte-nance
	(n)	%						
1	9037	14	6	23	16	35	14	21
2	4683	27	8	25	19	40	11	24
3	2650	36	8	28	23	44	9	26
4	1365	38	10	35	30	52	6	29
5	751	23	10	46	36	55	5	31
	% contact							
1	47	56	46	44	37	40	56	41
2	41	52	43	35	30	35	52	36
3	45	52	47	33	29	36	61	38
4	45	57	44	28	33	42	57	41
5	31	46	38	27	29	30	50	35
	β							
1	–0.064	0.517	0.041	0.030	–0.330	–0.228	0.278	–0.195
2	–0.324	0.741	0.269	–0.149	–0.532	–0.198	0.265	–0.196
3	0.033	0.512	0.411	–0.383	–0.700	–0.280	0.362	–0.204
4	0.183	0.834	0.212	–0.890	–0.779	0.153	0.193	–0.326
5	–0.634	0.949	0.516	–0.207	–0.261	–0.030	0.697	–0.034
	exp β							
1	0.938	1.678	1.042	1.030	0.719	0.796	1.320	0.823
2	0.723	2.098	1.309	0.862	0.587	0.820	1.303	0.822
3	1.034	1.669	1.508	0.682	0.497	0.756	1.437	0.815
4	1.201	2.302	1.236	0.411	0.459	1.166	1.213	0.722
5	0.531	2.584	1.676	0.813	0.770	0.970	2.008	0.967
	Wald							
1	3.61	67.32	0.20	0.33	25.69	20.72	19.11	13.18
2	41.03	113.91	6.02	4.06	36.49	8.13	7.53	6.91
3	0.22	35.51	7.36	15.96	41.36	9.41	6.42	4.48
4	2.68	44.64	1.10	46.24	29.29	1.28	0.61	5.93
5	15.43	24.31	3.81	1.45	1.93	0.02	3.36	0.04
	significance							
1	0.057	0.000	0.653	0.565	0.000	0.000	0.000	0.000
2	0.000	0.000	0.014	0.044	0.000	0.004	0.006	0.009
3	0.642	0.000	0.007	0.000	0.000	0.002	0.011	0.034
4	0.102	0.000	0.294	0.000	0.000	0.259	0.435	0.015
5	0.000	0.000	0.051	0.229	0.165	0.875	0.067	0.848

Table .4 (cont.) Logistic regression contact rate per call on households not previously contacted; eligible households[a]

number call	total/ constant	evening call	Saturday call	no listed phone number	msterdam, Rotterdam The ague	highrise buildings	detached dwelling	poor mainte-nance
	(n)	%						
	standard error							
1	0.034	0.063	0.092	0.052	0.065	0.050	0.064	0.054
2	0.051	0.069	0.110	0.074	0.088	0.069	0.096	0.075
3	0.072	0.086	0.151	0.096	0.109	0.091	0.143	0.096
4	0.112	0.125	0.202	0.131	0.144	0.136	0.248	0.134
5	0.161	0.193	0.264	0.172	0.188	0.191	0.380	0.178
	$\chi 2$ (df 7)		Cox & Snell R^2		Nagel-kerke R^2			
1	203		0.022		0.030			
2	214		0.045		0.060			
3	173		0.065		0.086			
4	153		0.113		0.150			
5	36		0.053		0.072			

a Calls for which timing visits is unknown are excluded, betas of reference categories are 0. Dependent is contact(yes)

From the analyses above it appears that contactability is related to the timing of the calls, the possibility of telephoning the sample household and urbanicity. It could be assumed, as do Groves and Couper (1998, section 4.6) that the effect of urbanicity is due to differences in household composition and '... higher frequencies of being away from home in urban areas for reasons of commuting to work, recreation, and other opportunities'. Whether or not this is true should emerge from the analysis in the next section of the contactability of the respondents, about whom we have information concerning these aspects.

7.3.2 *Contactability responding households*

Figure 7.6 illustrates the contactability of different types of households and individuals in the same way as in figure 7.5. At each subsequent call to previously non-contacted households the composition of the remaining noncontacted sample has changed because the contacted households are excluded and these (relatively) easy-to-contact households differ from those for whom more calls are required. Figure 7.6 includes only those households that responded, and as a consequence comprises more substantive information on these households. Therefore, information has been included on a number of correlates of contactability that were identified in section 3.2, namely single-adult households, the presence of children and being employed or a student. Going to the cinema also appeared to be related to contactability.

Fig re 7.6a Number of calls to first contact and percentage big city (above), detached dwelling (middle) and young single households (below) (respondents)

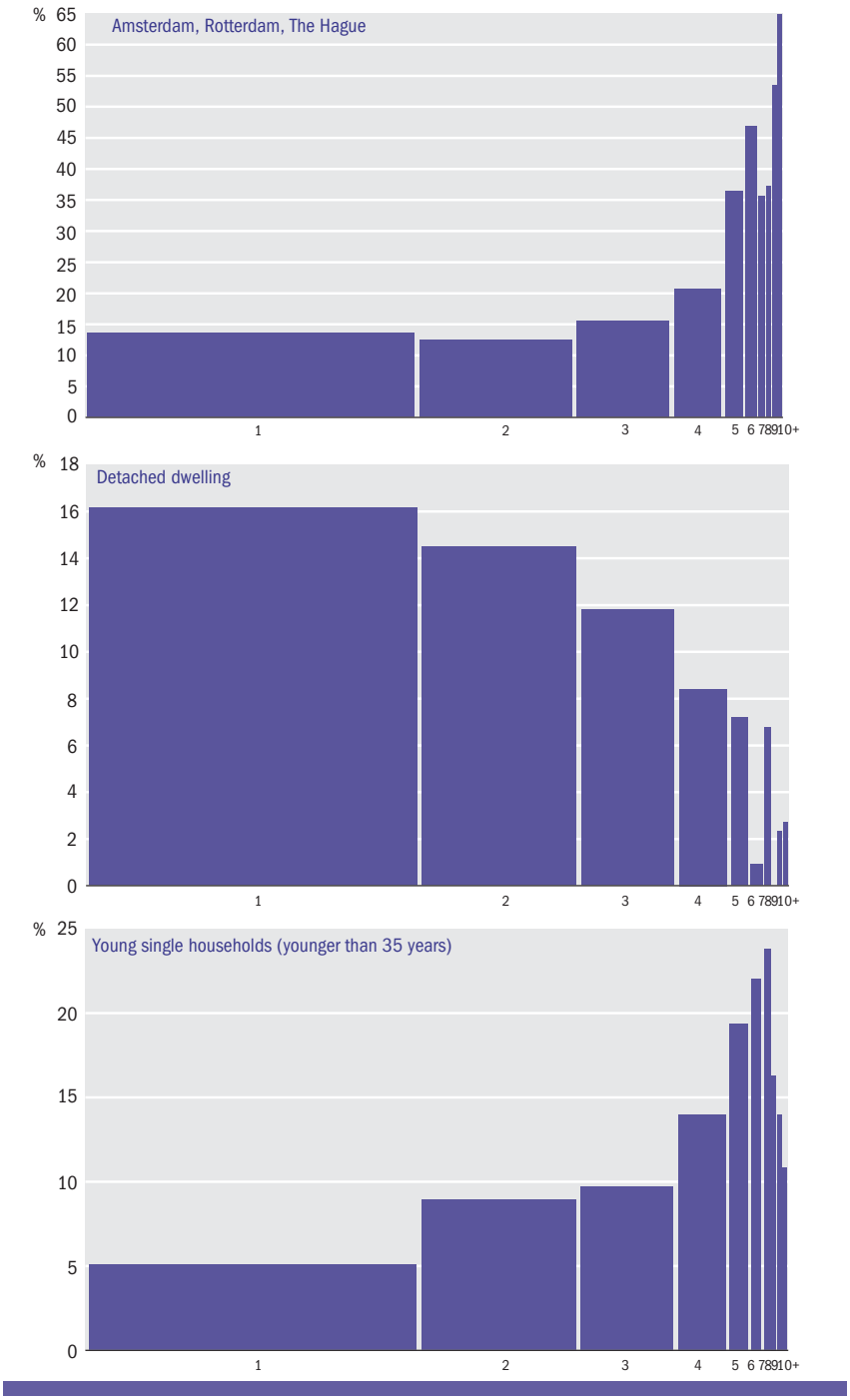
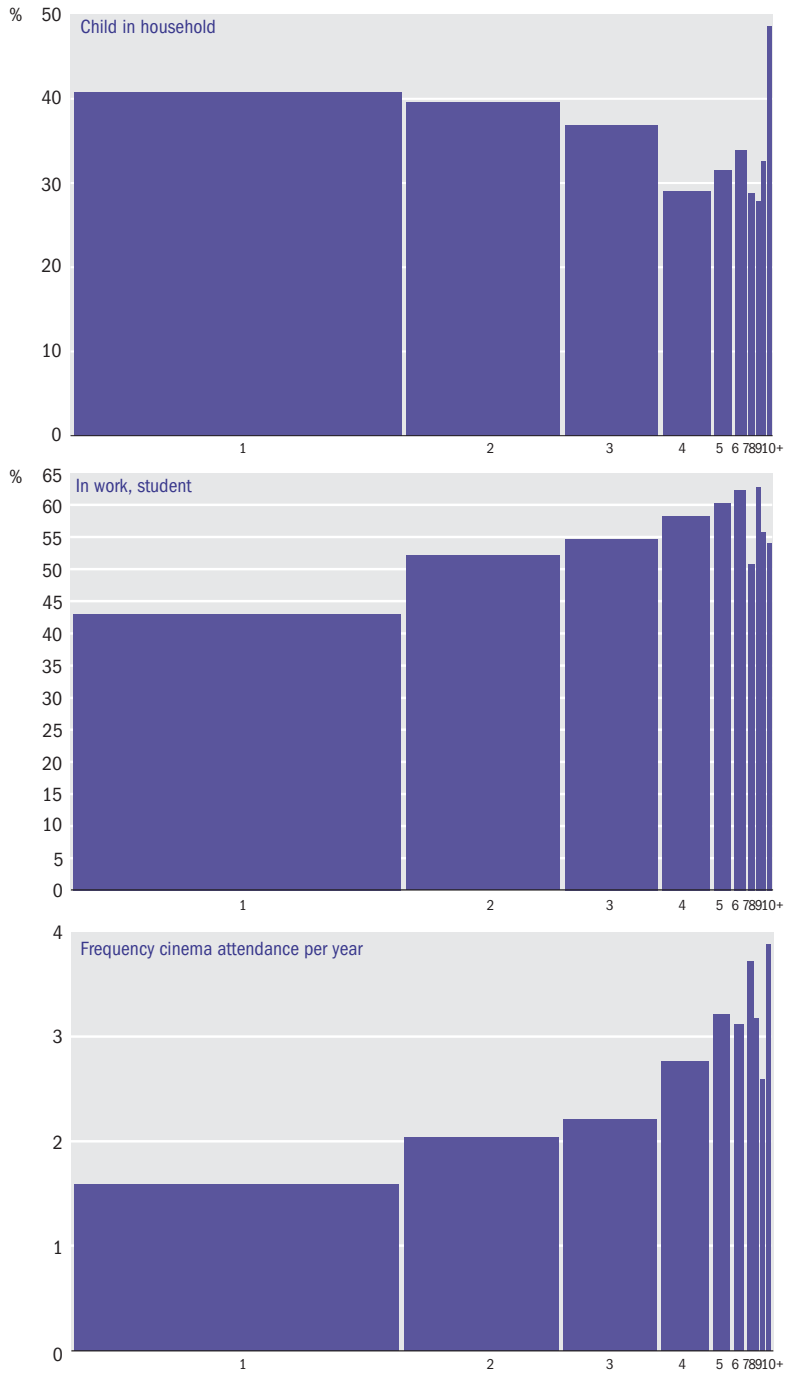


Figure 7.6b Call to first contact and percentage child in household, having a job or being a student and frequency cinema attendance (respondents)



The pattern for urbanicity (big cities and the individual indicator 'detached dwelling') is similar to that for the eligible households. Figure 7.6 shows that young single-adult households are hard to reach. The share of singles aged 34 years or younger increases up to the seventh call, but then decreases for those households (or individuals in this case) that take more than seven calls to first contact. The same holds for having children, being in work or a student, and being a frequent cinema-goer. In all cases the expected relationship holds except for the very last calls. As only a very small proportion of the respondents could not be reached until these final calls (about 150 responding households required more than seven calls to first contact) it cannot be determined whether this is a systematic effect, or perhaps just a late correction for deficient fieldwork in a rural area. In the more systematic analyses in table 7.6 the focus will again be on the first five calls.

Table 7.5 presents statistics and the results of logistic regressions of the likelihood of making contact among previously noncontacted responding households, following the same format and introducing the same dependent variable 'contact (yes)' as in table 7.4. Independent variables are whether or not the call was made in the evening, the availability of an unlisted phone number, big-city residency, poor maintenance of the dwelling/neighbourhood, living in a detached dwelling, having a child at home and two indicators of the daily activities of the household and respondent: the number of persons in paid employment in the household, and whether the respondent considers his or her main occupation to be 'employed' or 'student'. The first variable, with the categories 'single-person household, not in paid employment', 'single-person household, in paid employment', and 'multiple-person household', subdivided into 'no-one in paid employment', 'single-earner household' and 'two-earner household' should reflect the likelihood of a member of the household being at home during working hours, whereas the second focuses on the social position and the at-home pattern of the respondents with whom contact was made.

Finally, the type of cultural participation and age were included. Age was included because section 3.2 clearly identifies older people as being easier to contact. Rather than including 'going to the cinema' as an independent variable, as was done for illustrative purposes in figure 7.6, in the analyses in table 7.5 a more global indicator of cultural participation is included as an indicator of both lifestyle and an activity that takes people out of the house. The variable distinguishes between those who participate in popular culture (cinema, dance events, pop concerts), classical culture (theatre, concerts, opera, etc.), both types and no participation at all.

In table 7.5 only those variables are included that were expected to be related to ease of contact from earlier studies and that had a significant relationship with contactability. Sex, which was strongly related to being at home in the time use study (see figure 3.1), had no effect at all. This is most probably due to the fact that the AVO is a household sample and the initial contact could be made with any knowledgeable adult. Several substantive variables that were assumed to be related to contactability,

such as membership of organizations and sports activities, are not included in table 7.5 because they had no effect at all, although the frequency of watching sports on TV had a negative effect on contactability at the first call and a positive effect at subsequent calls. There does not seem to be an explanation for the latter result.

The first five rows show the sample composition across calls. Excluding respondents on whom no call records were available, the table shows that 5898 respondents received a first call. 14% of these first calls were in the evening and 11% of these respondents were single persons with a job. 3045 respondents received a second call (which means that 5898-3045 = 2853 of the respondents were contacted at the first call. The percentage of evening calls increases up to the fourth call and then drops at the fifth, which is likely to be related to the wave structure of the fieldwork discussed in section 6.2.2. As for all sample persons, the percentage of unlisted phone numbers, of big-city dwellers, and of those living in poorly maintained neighbourhoods and dwellings increases with the number of calls required to make contact. Living in a detached dwelling and having a child at home decreases with the number of calls. Striking with regard to household labour market participation is the increasing percentage of single persons with a job, namely 11% at the first call and 21% at the fifth call.

The percentage of individuals among the respondents who are employed/students increases from 49 % to 59% and the percentage of those who do not partake in cultural activities decreases from 30% of those who were contacted at the first call to 24% at the fifth call. The age of early contacted respondents is 49.3 years and of late contacted respondents (fifth call) 46.1.

The second block of five rows in table 7.5 represents the contactability of each category of respondents at each subsequent call. As in table 7.5, a lower contactability at a particular call of a particular group will be reflected by a larger share of this group at the next call. There is no reason, however, why the contact rate of a particular group should differ between calls except when different fieldwork approaches are used between calls. The table shows that the contact rate within groups differs between calls but that there is no clear pattern.

Next in table 7.5 the outcomes of five separate logistic regressions are presented, for the five first calls to previously noncontacted households. One should keep in mind that these outcomes are based only on those households that cooperated in the survey. As in table 7.4, each block of five rows comprises outcome parameters for each of these five regressions. The results for the call pattern and socio-structural characteristics are very similar to table 7.4. Evening calls had a large positive effect on contactability, having an unlisted telephone number a negative effect for call 2, 3 and especially 4. Inhabitants of big cities and those living in poorly maintained neighbourhoods and dwellings were harder to contact, and the residents of detached dwellings easier. The presence of children at home also had a positive effect for the first two calls.

Table .5 Logistic regressions for call 1 to 5 to previously noncontacted households[a]

								persons with jobs						cultural participation			
								single person		more persons, # earners							
call number	**total/ constant**	**evening call**	**unlisted phone number**	**big cities**	**poor maintenance**	**detached dwelling**	**child at home**	no job (ref.)	job	none	single	multiple	**employed student**	none (ref.)	both	classical	popular	**age**
	(n)	%																average
1	5898	14	21	16	20	14	38	16	11	18	22	33	49	30	31	22	17	49.3
2	3045	27	23	19	23	11	36	16	14	16	21	33	55	28	34	21	18	47.6
3	1672	37	26	24	27	9	34	16	17	15	20	32	57	26	34	21	19	46.8
4	833	38	33	33	32	6	31	18	20	14	19	29	58	24	35	20	21	46.4
5	449	22	42	43	36	4	33	18	21	14	19	27	59	24	35	19	21	46.1
	% contact																	
1	47	58	44	38	39	56	50	49	30	56	50	45	41	52	41	50	45	51
2	43	53	37	29	36	54	47	43	33	44	45	46	41	47	43	42	38	48
3	47	53	35	29	38	62	51	39	37	52	49	51	45	50	45	50	41	47
4	47	56	31	31	39	63	45	47	46	48	46	49	47	47	46	49	47	47
5	36	58	36	31	36	61	35	41	34	43	38	30	36	40	41	28	32	47
	β																	
1	-0.622	0.593	0.007	-0.310	-0.304	0.263	0.301		-0.222	0.216	0.086	0.070	-0.233		-0.180	-0.101	-0.003	0.010
2	-0.143	0.750	-0.210	-0.733	-0.229	0.320	0.211		-0.254	-0.083	-0.076	0.052	-0.230		-0.060	-0.244	-0.300	0.001
3	0.083	0.501	-0.497	-0.866	-0.203	0.331	0.117		0.193	0.401	0.281	0.485	-0.329		-0.135	-0.001	-0.316	-0.002
4	0.742	0.605	-0.822	-0.953	-0.298	0.350	-0.228		-0.008	-0.009	-0.191	0.033	-0.178		-0.070	-0.024	-0.014	-0.004
5	-1.055	1.203	-0.013	-0.321	-0.025	1.000	0.166		-0.738	0.005	-0.408	-1.057	0.665		0.159	-0.645	-0.301	0.011

Table .5 Logistic regressions for call 1 to 5 to previously noncontacted households[a] (continuation)

								persons with jobs						cultural participation			
								single person		more persons, # earners							
call number	total/ constant	evening call	unlisted phone number	big cities	poor maintenance	detached dwelling	child at home	no job (ref.)	job	none	single	multiple	employed student	none (ref.)	both	classical	popular	age
	exp. β																	
1	0.537	1.809	1.007	0.734	0.738	1.301	1.351		0.801	1.241	1.090	1.073	0.792		0.835	0.903	0.997	1.010
2	0.867	2.117	0.811	0.481	0.796	1.377	1.235		0.775	0.921	0.927	1.054	0.794		0.941	0.783	0.741	1.001
3	1.087	1.651	0.608	0.421	0.816	1.393	1.125		1.212	1.493	1.324	1.625	0.720		0.874	0.999	0.729	0.998
4	2.099	1.832	0.439	0.386	0.743	1.420	0.796		0.992	0.991	0.826	1.034	0.837		0.933	0.976	0.986	0.996
5	0.348	3.330	0.987	0.726	0.976	2.719	1.181		0.478	1.005	0.665	0.347	1.944		1.173	0.525	0.740	1.011
	Wald																	
1	11.03	56.75	0.01	16.55	18.87	10.80	18.10		2.75	5.51	0.65	0.36	8.05		5.85	1.86	0.00	18.00
2	0.30	76.09	4.83	48.35	5.84	6.95	4.54		2.07	0.37	0.25	0.10	3.88		0.34	4.96	5.65	0.13
3	0.06	21.62	16.08	44.70	2.73	3.27	0.74		0.74	4.61	1.82	4.88	4.21		0.90	0.00	3.50	0.25
4	2.17	15.27	24.64	32.57	3.12	1.18	1.19		0.00	0.00	0.41	0.01	0.56		0.11	0.01	0.00	0.47
5	2.26	24.35	0.00	2.11	0.01	3.48	0.30		3.16	0.00	0.94	5.73	3.57		0.32	3.80	0.80	1.35
	significance																	
1	0.001	0.000	0.920	0.000	0.000	0.001	0.000	(df 4)	0.097	0.019	0.421	0.549	0.005	(df 3)	0.016	0.173	0.972	0.000
2	0.585	0.000	0.028	0.000	0.016	0.008	0.033		0.150	0.541	0.619	0.748	0.049		0.561	0.026	0.018	0.721
3	0.811	0.000	0.000	0.000	0.098	0.070	0.391		0.390	0.032	0.177	0.027	0.040		0.344	0.993	0.061	0.615
4	0.141	0.000	0.000	0.000	0.077	0.278	0.274		0.978	0.974	0.521	0.915	0.456		0.737	0.914	0.952	0.492
5	0.132	0.000	0.952	0.146	0.913	0.062	0.586		0.076	0.988	0.332	0.017	0.059		0.571	0.051	0.370	0.245

Table .5 Logistic regressions for call 1 to 5 to previously noncontacted households[a] (continuation)

call number	total/ constant	evening call	unlisted phone number	big cities	poor maintenance	detached dwelling	child at home	persons with jobs: single person: no job (ref.)	persons with jobs: single person: job	persons with jobs: more persons, # earners: none	persons with jobs: more persons, # earners: single	persons with jobs: more persons, # earners: multiple	employed student	cultural participation: none (ref.)	cultural participation: both	cultural participation: classical	cultural participation: popular	age
	standard error																	
1	0.187	0.079	0.069	0.076	0.070	0.080	0.071		0.134	0.092	0.107	0.118	0.082		0.074	0.075	0.089	0.002
2	0.262	0.086	0.096	0.105	0.095	0.121	0.099		0.177	0.135	0.153	0.163	0.117		0.104	0.110	0.126	0.003
3	0.348	0.108	0.124	0.129	0.123	0.183	0.137		0.224	0.187	0.208	0.220	0.160		0.143	0.152	0.169	0.005
4	0.504	0.155	0.166	0.167	0.169	0.323	0.209		0.301	0.269	0.297	0.312	0.239		0.208	0.226	0.238	0.007
5	0.701	0.244	0.218	0.220	0.226	0.536	0.305		0.415	0.368	0.420	0.442	0.352		0.281	0.331	0.336	0.009

	X^2 (df 15)	Cox & Snell R^2	Nagelkerke R^2
1	296	0.049	0.065
2	190	0.061	0.081
3	144	0.083	0.110
4	101	0.115	0.153
5	49	0.104	0.143

a Dependent variable: contact(yes). β reference categories 0.

Surprisingly, the labour-market position of the household seems to be less relevant for contactability than the respondent being employed or a student. The latter characteristic had a negative impact on contactability at the first three calls. The β values for 'persons with jobs' are much smaller and have positive and negative effects across calls. The analyses also show that those who participate in cultural activities, irrespective of the type, are less homebound than nonparticipants. As De Haan & Knulst (2000) and De Haan (2001) have shown, city-dwellers, singles and the higher educated are culturally more active in the Netherlands. Here we see that being a city-dweller, being single (and in paid employment) and being culturally active are cumulative factors in decreasing the likelihood of contact. Age, the last column in table 7.5, is related to contactability, as might be expected from the time use data in figure 3.1, although only for the first call.

Campanelli, Sturgis and Purdon (1997, section 3.4) modelled the probability of contact at a particular call, assuming noncontact prior to that call, as a discrete hazard rate logistic regression model (see annex 7.1) in which the probability is estimated that contact will be made at call **t** given that no contact has occurred before call **t** or – put differently – the probability that, if an interviewer calls at an address at a particular time, in a particular area, and at a particular time of year, the door will be answered. In their final model they included the following variables: the time of the current call and of the immediately preceding call, the number of previous calls, the month, the nature of the area and the interviewer grade. The upper part of table 7.6 presents the results of a similar analysis, including the time of the current call, the number of previous calls, the nature of the area (big city, high-rise buildings), and the availability of a listed telephone number.

Table .6 iscrete hazard rate logistic regression model for contacts (dependent: contact yes)[a]

		exp( )	Wald	sig.	S.E.
calls on previously noncontacted eligible sample households[b]					
unlisted telephone number	–0.334	0.716	101.471	0.000	0.033
evening call	0.504	1.655	212.555	0.000	0.035
Amsterdam, Rotterdam, The Hague	–0.348	0.706	76.778	0.000	0.040
high-rise buildings	–0.202	0.817	39.182	0.000	0.032
detached dwelling	0.309	1.362	44.985	0.000	0.046
number of previous calls	–0.101	0.904	139.618	0.000	0.009
constant	–0.018	0.982	0.594	0.441	0.024
calls on previously noncontacted respondents[c]					
unlisted telephone number	–0.213	0.808	23.921	0.000	0.044
evening call	0.528	1.695	146.572	0.000	0.044
Amsterdam, Rotterdam, The Hague	–0.479	0.620	94.843	0.000	0.049
high-rise buildings	–0.116	0.890	7.051	0.008	0.044
poor maintenance	–0.210	0.811	22.239	0.000	0.044
detached dwelling	0.268	1.307	20.505	0.000	0.059
child	0.151	1.163	10.046	0.002	0.048
age	0.004	1.004	5.703	0.017	0.002
persons with jobs (df 4)			11.842	0.019	
single person, no job (ref.)					
single person, job	–0.126	0.881	2.336	0.126	0.083
more persons, no earner	0.125	1.133	3.880	0.049	0.064
more persons, single earner	0.014	1.014	0.036	0.850	0.072
more persons, multiple earners	0.044	1.045	0.321	0.571	0.078
employed, student	–0.179	0.836	10.343	0.001	0.056
cultural participation (df 3)			6.579	0.087	
both classical and popular	–0.103	0.902	4.326	0.038	0.049
classical	–0.107	0.898	4.390	0.036	0.051
popular	–0.109	0.897	3.445	0.063	0.059
none (ref.)					
constant	–0.136	0.873	1.166	0.280	0.126

a See annex 7.1 for a description of this technique.
b Chi^2 856 (df 6), Cox & Snell R^2 0.040 Nagelkerke R^2 0.053 N 21,034
c Chi^2 590 (df 16), Cox & Snell R^2 0.044 Nagelkerke R^2 0.058 N 13,209

The timing of the preceding call has not been included in the analysis as it did not significantly improve the prediction of contact rates (nor did it in the analysis by Campanelli et al.). Month was excluded because early first contacts are by definition near the start of fieldwork in September (see annex 6.2). Interviewer characteristics were not included as these were missing for particular parts of the country (see annex 6.3). As an additional predictor of contact, or rather noncontact, the absence of a listed phone number was included in the model. As was found by Campanelli et al., the odds of making contact are high for evenings and low in urban areas. House-

holds without a listed phone number are also harder to reach. Contact becomes less likely as the number of calls increases (every extra call is associated with a 10 percent reduction in the odds of contact).

The lower part of table 7.6 contains the results of the same model for respondents only. Here the effect of the number of previous calls was 0. The results are very similar to those of the first call in table 7.5, although in that analysis there was no effect for the availability of a listed phone number and the contribution of the background variables was higher. This might have been expected, as the telephone started to play a role in later calls only, and the importance of background characteristics seemed to diminish at subsequent calls, as shown in table 7.5.

7.3.3 Continuum of resistance and the effect of an increased number of calls

In section 5.2 the concept of a 'continuum of resistance' has been introduced. According to Lin and Schaeffer (1995, p. 238) this method for assessing the impact of nonparticipation on survey estimates '... assumes that nonparticipants are all similar to each other and that if nonparticipants took part in the study, they would respond like a particular group of participants – usually those participants who were difficult to contact for the interview'. They investigated whether there is a linear relationship between contactability and the amount of child support, and whether nonparticipants owe and pay a similar amount of child support to late contacts. In their study child support was closely related to the topic and was assumed to be related to contactability. The variable was available for each sample person.

As the present study comprised hardly any noncontacts (see section 6.2.2), the investigation will focus on degrees of contactability and ignore the question of whether those who were very hard to reach are similar to the final noncontacts. In addition, little information is available on refusals, except for the small subsample that participated in the follow-up survey. For this reason, the existence of a continuum of resistance will be investigated for respondents only. This does not seem too much of a limitation as the fact remains for these respondents that with a smaller number of calls a proportion of them would have been nonparticipants. Figure 7.5 showed that there is a relationship between survey variables and number of calls to first contact that is ordinal for the first call but not for the latter. Table 7.7 presents similar results. The table shows that up to the sixth call to first contact participation in activities that take one away from home (culture, sports, work and study) decreases, but also that those respondents that were contacted after the sixth call show an erratic pattern. The outcomes are not at all similar to the ideal model presented in figure 5.1b. They are however far less irregular than those presented by Lin and Schaeffer, at least up to the sixth call.

The bumpy pattern at the end of the continuum might be explained by the dual nature of noncontact. Schnell (1997) observed that persons may be hard to catch at home either because they are not at home at usual hours – i.e. hours that are convenient for interviewers – or because they are away for prolonged periods (see section 2.6.2). It might well be that six calls suffice to catch the first group at home, provided the calls are spread over different times of the day and days of the week, and even more so if telephone calls can be made. Later calls might finally lead to contact with respondents who were temporarily away and came back when the fieldwork was well under way. This dual nature of contactability undermines the idea of a continuum of resistance. Without prolonged absences of sample persons, and ignoring the households that require more than six calls, the existence of a weak continuum of resistance might be sustainable.

Table 7.7 also presents the correlations between not-at-home activities and number of calls to first contact, which are rather small but significant. In addition, it introduces age and urbanicity as background variables that are ordinally related to the number of calls to first contact up to the sixth call, which is in line with the idea that difficulty of contact has both a short-term and a long-term component. The correlation of calls to first contact with age and urbanicity is much higher than with the participation variables. The table also shows the correlations between background variables and not-at-home activities. And finally, the correlation between calls to first contact and the participation variables decreases substantially after taking urbanicity, and especially age into account. This suggests that in the early stages subsequent calls to previously noncontacted households increasingly recruit younger (urban) respondents who participate more in not-at-home activities. The strong relationship with age suggests that this could be a case of Missing At Random (ONS) rather than Not Missing at Random (NMAR) (see section 2.3). The effect of noncontact had fewer calls been made might be adjusted for by weighting for age. The outcomes of the analyses presented in table 7.4 and 7.5 suggest, however, that background variables would not be entirely sufficient in adjusting for noncontact bias – had fewer calls been made – with respect to cultural participation and work or study. In summary, the conclusion might be that there seems to be a weak continuum of resistance that may partly be concealed by respondents who were not at home during the first part of fieldwork and that is partly – but not entirely – a consequence of the different socio-demographic composition of early and late respondents.

After having decided on a 'light' continuum of resistance, the next question to be answered in this section is whether increasing the contact rate improves survey quality (see section 5.3). Contacting those who are hard to contact means either more calls or calls at times when households are more likely to be at home, or calls at a later stage when the household may have returned from a prolonged absence. The tables and figures presented in this chapter so far indicate that a large number of calls to previously noncontacted households can substantially increase response rates, change the composition of the response group and possibly diminish nonre

Table . ultural and social participation by number of calls to first contact

	culture %		sports		job or study		urbanicity[a]		age		N
1	67		58		43		3.04		51.3		2941
2	70		61		52		3.04		48.2		1399
3	73		61		55		2.80		47.7		863
4	77		62		58		2.66		46.8		450
5	75		66		60		2.30		46.6		181
6	81		64		62		2.02		44.4		109
7	75		59		51		2.42		46.8		59
8	77		63		63		1.91		44.0		43
9	72		70		56		2.02		45.6		43
10	69		69		55		1.52		48.4		29
11	75		88		50		2.25		53.9		8
total	70		60		49		2.91		49.3		6125
R(calls to first contact)	0.065	**	0.041	**	0.100	**	-0.194	**	-0.180	**	
R(age)	-0.269	**	-0.306	**	-0.591	**	0.054	**			
R(urbanicity)	-0.073	**	-0.016		-0.058	**					
R(calls to first contact|age)	0.041	**	0.012		0.055	**					
R(calls to first contact|urbanicity)	0.052	**	0.038	**	0.092	**					
R(calls to first contact|age,urbanicity)	0.030	*	0.012		0.050	**					

a Urbanicity: Number of inhabitants per square kilometre, 5 classes, from very urban to not at all urban.
** Significant at .01; * Significant at .05

sponse bias. The question of whether an increasing number of calls, and the resultant higher contact rate, improve the survey quality can be answered by introducing independent evidence. In figure 7.7a the composition with respect to urbanicity, an important correlate of contactability according to the previous section, of the contacted sample after each subsequent call to first contact is compared to the composition of the net eligible sample (see figure 2.2). This might suffer from sampling errors but these have nothing to do with nonresponse. Figure 7.7a illustrates again how big city-dwellers profit from each subsequent call. After each additional call the contacted sample becomes more similar to the net sample.

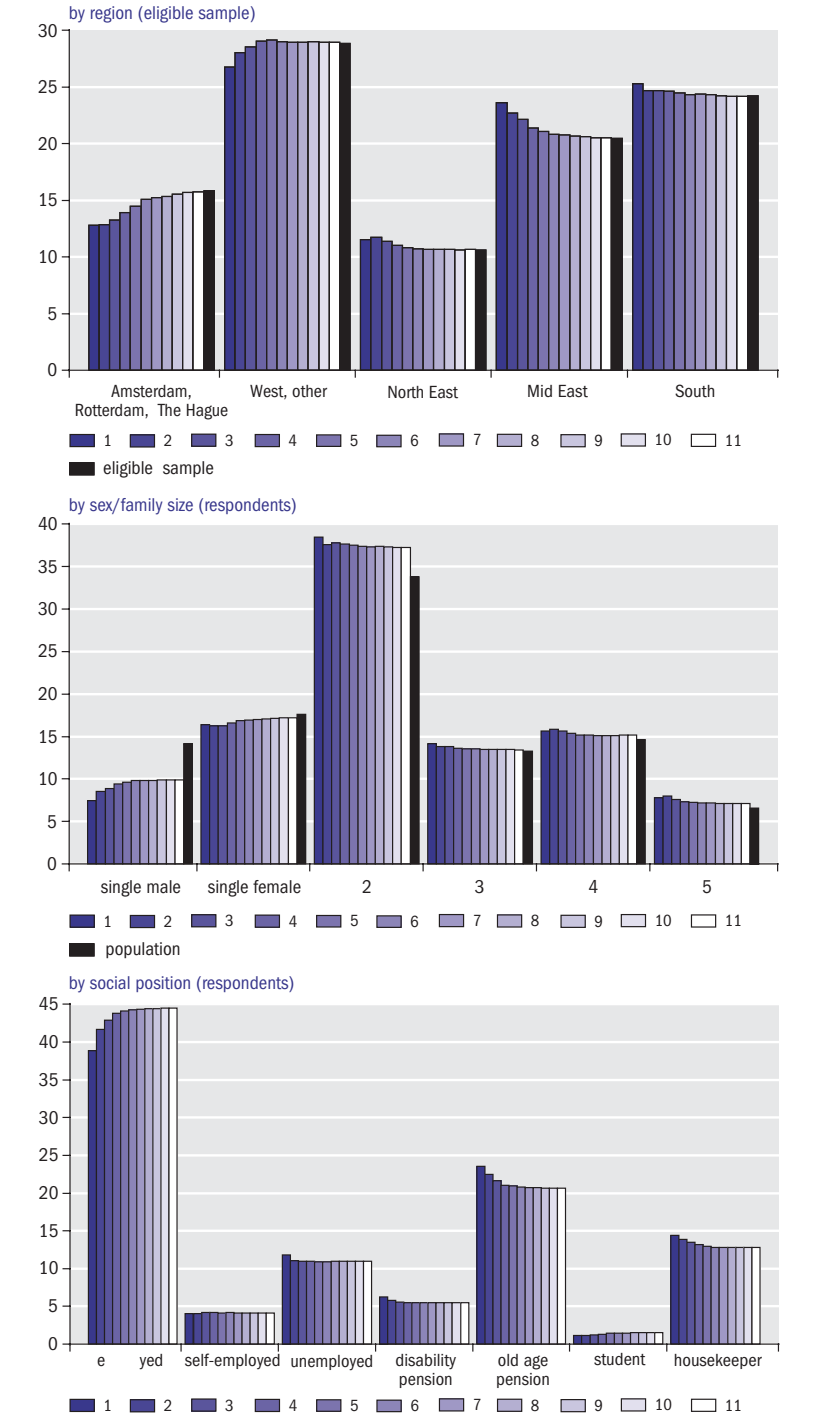
Figure 7.7b shows the effect of increasing the number of calls on family size. No information was available on the family size of the net sample, but only on the responding households. Independent evidence is introduced by comparing the family size of the households contacted at each additional call with population data from Statistics Netherlands. Population totals have been corrected for the number of institutional

households, assuming that these will solely consist of singles and equally of men and women (CBS 2000; 2001). This comparison might confound sampling error with nonresponse error. The figure shows that the distribution according to family size becomes more like the population with each call. Single males, however, remain vastly underrepresented, two person households slightly overrepresented.

Figure 7.7c illustrates that employed persons and students are better represented after each subsequent call, although the influence of later calls is very small. In this case, neither a comparison with the net sample nor with the target population was available, as the social position of the net sample is unknown and no population data are available on the selected knowledgeable adult per household. As those who are employed or follow a study have been identified as hard to contact in the literature presented in section 3.2, it can be assumed that establishing contact with this group will improve survey quality and reduce nonresponse bias.

Based on the evidence presented in this section , the conclusion is that increasing the contact rate improves survey quality, at least up to the sixth call. Rather than making many calls within a short period, it might be more effective to make up to six calls, spread across times of the day and days of the week, within a relatively short period and then re-approach noncontacted households after some time in case they have been away for a prolonged period. This of course requires a longer fieldwork period.

Figure 7.7 Number of calls to first contact



7.4 Total number of calls: contact and cooperation

Through lack of detailed call records many nonresponse studies are based on a single indicator of 'difficulty', namely the total number of calls (see section 2.6.4). In this study a better – although still imperfect – measure could be used, namely the number of calls to first contact. Distinguishing calls to establish contact and further calls to obtain cooperation separates the two main factors in survey participation. Table 7.8 presents the correlation between different measures of field efforts in the AVO.

Table . orrelation between indicators of contactability and cooperation

	contacted sample households	respondents	early contacts (1-4)
R (number of calls ; times not at home)	0.91	0.91	0.87
R (number of calls ; number of calls to first contact)	0.78	0.72	0.55
R (number of calls ; number of contacts)	0.45	0.48	0.69
R (number of calls to first contact; number of contacts)	-0.16	-0.16	-0.06
R (number of calls to first contact; number of unsuccessful contacts)	-0.09	-0.09	0.00
N	9261	6125	5653

As might be expected, there is a high correlation between the total number of calls and the times when nobody is found at home, for all eligible sample households, all respondents and also for those households that were moderately easy to contact and required at most four calls to first contact. The correlation between the total number of calls and the number of calls to first contact is considerably lower and between the total number of calls and the number of contacts (interview, appointment, broken off, not able, refusal) still lower. The correlation between the best indicator of contactability (number of calls to first contact) and indicators of cooperation (number of contacts, number of unsuccessful contacts (not able, refusal)) is very small and negative. For the easier contacts, this correlation is close to zero, which suggests that for more difficult-to-contact households there may have been fewer occasions for subsequent contacts.

Figure 7.8 shows a further illustration of the relationship between contactability and cooperation. Figure 7.8a depicts the outcome of the call at which first contact was established, which is not necessarily the final outcome. If contact is made at the fourth call, the percentage of appointments is particularly small, which will at least partly be due to non-recorded first contacts by telephone that resulted in an interview at the fourth call. Also notable is the higher percentage of (temporary) refusals at the fourth call, which might be related to the fourth call being the last one of the first wave and which might be a refusal on the phone, and the larger percentage of

Figure 7.8a Outcome call during hich first contact as made by number of call

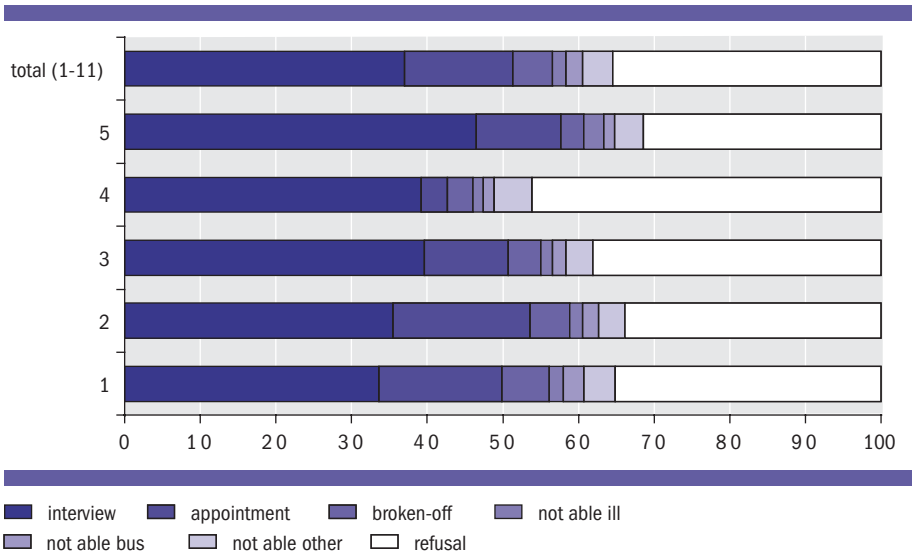
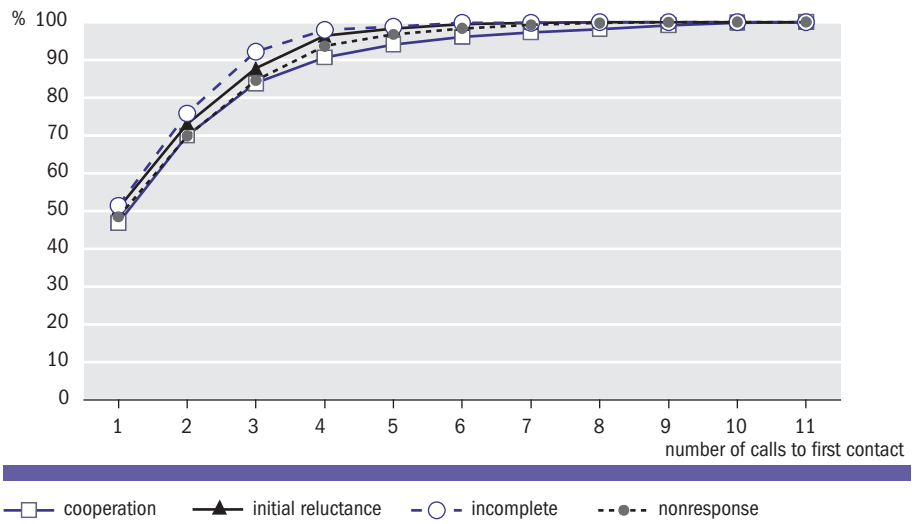


Figure 7.8b Number of calls to first contact by final outcome



interviews at the fifth call. Figure 7.8a gives no clear indication that sample households were less cooperative if many calls were needed to make contact, and nor does figure 7.8b, which presents the cumulative distribution of the number of calls required to make contact by the final outcome of all calls. Households that ultimately returned incomplete questionnaire sets, and households that were initially reluctant to cooperate (outcome first contact 'not able' or 'not willing') seem to require slightly fewer calls to first contact than either the cooperative households or the final nonrespondents. The differences are very small, however.

Figure 7.9 Number of contacts (1-5+) for number of calls to first contact (1-10+)

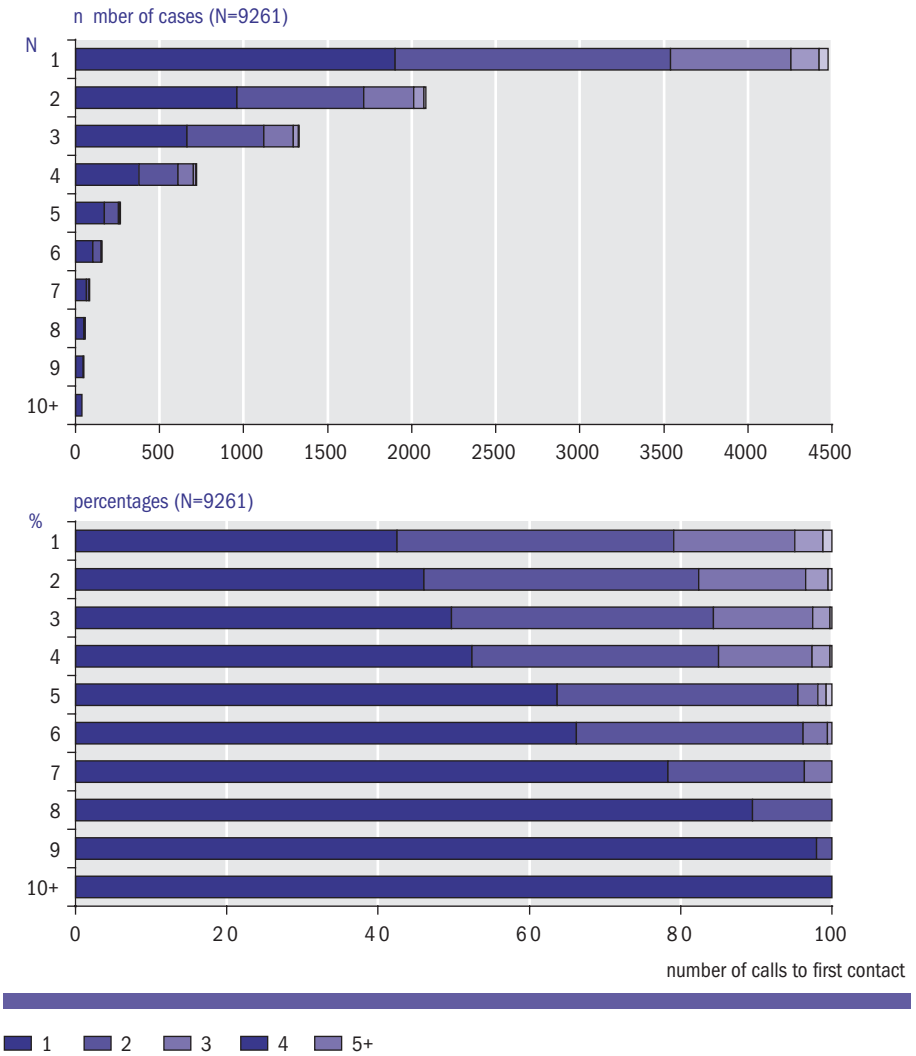


Figure 7.9 depicts the relationship between contactability and reluctance in a different way. It presents the number of contacts for the households that were first contacted at the first call, at the second call, etc. As the absolute figures in the upper part of the figure show, the number of first contacts decreases rapidly after the first call. The lower percentage section shows that first contacts at later calls are less likely to be succeeded by additional contacts. As mentioned earlier, these later contacts are scarce.

All evidence presented in this section seems to confirm that late contacts are not necessarily less likely to cooperate in a survey. If sample units are contacted at later stages, however, they may participate less frequently, simply because there is less time available for making additional calls on possibly temporary refusals, and thus for refusal conversion.

7.5 Conclusions

The results of the analyses of contactability from the AVO1999 study presented in this chapter are both straightforward and complicated. Straightforward is the outcome that households in urban areas, persons who pursue activities outside the home and persons who are part of smaller households are more difficult to catch at home. This could be expected from the literature on the process of establishing contact presented in section 2.6.2 and on who is generally hard to contact, presented in section 3.2. The following groups have been identified as requiring many contact attempts: persons in urban areas, sample persons who have activities outside the home and sample persons who are members of smaller households. A complicating factor mentioned in section 7.2 was that interviewers may adapt their calling patterns to neighbourhood or dwelling characteristics of which they are aware through local knowledge or previous visits.

Section 7.2 also confirmed the results from earlier studies, that the availability of a listed telephone number is unevenly distributed across the Dutch population, which will make specific groups, e.g. urban dwellers, ethnic minorities, younger persons, harder to reach if telephone recruitment is allowed. These very same groups are also less likely to be caught at home. This implies that telephoning sample persons who were not found at home may not reduce noncontact bias, as the groups that cannot be reached by telephone share their background characteristics with those who are less housebound. The AVO strategy of telephoning noncontacted households at the end of each wave takes this possible bias into account, as those who could not be reached by telephone were called on in person again in the next wave.

The combined results on number of calls and telephone accessibility thus clearly identify the groups that are hard to contact. The results from the preceding sections also give guidance on how to enhance contact rates and how to improve call efficiency, at least if efficiency is defined as the number of calls required to established first contact. The unequivocal answer should be that evening calls have the highest contact rates (figure 7.1) and that, as cooperation is independent of the timing of the call (section 7.2), making evening calls minimizes the number of contact attempts. However - and this is where things start to become complicated - it was also stated in section 7.2 that restricting contact attempts to the evening might increase the contact rate of individual calls but might also increase fieldwork duration, as only a limited number of calls can be made during the evening, and fieldwork costs, as

evening calls might require more travel and may be more expensive. Efficient call schedules are an intrinsic part of telephone interviewing but may be difficult to implement in face-to-face interviewing, if only because interviewers seem to prefer to have some leeway in organising their calls and may base this on personal preferences and local knowledge.

From the literature and the current analyses the following procedures are recommended:
- Spread call attempts across times of the day and days of the week;
- Ensure that one of the first calls to previously noncontacted households is an evening call;
- If telephone calls are made to previously noncontacted households, make sure that households without a listed telephone number are re-approached too;
- Because of possibly longer absences, make sure that noncontacted households are re-approached after a few weeks;
- Ensure that independent information is available to distinguish between ineligible sample units (empty dwellings, holiday homes) and final noncontacts (nobody found at home);
- Record and key call information both to improve efficiency, monitor the calling process, analyse contactability and if possible correct for noncontact bias;
- Use the number of calls required to establish first contact as an indicator of contactability.

The empirical evidence presented in this chapter demonstrates that contact forms and detailed call records can provide the information required to analyse contactability and to improve fieldwork strategies. Despite the inadequacies of the current forms, they can be considered as a valuable tool to study nonresponse and to improve contact strategies. On the basis of the call data it could be confirmed that contactability and cooperation are separate nonresponse factors. It may be true that persons who are hard to contact and reluctant to cooperate share a number of the same characteristics. However, section 7.4 offers no evidence that persons or households who are first contacted at later calls are more reluctant to cooperate than those who are contacted earlier. Persons who are contacted at later calls may participate less because there may not be time left for additional calls at more favourable times or by a different interviewer.

Based on the information from the contact forms, the tentative conclusion is that in this study there is a weak continuum of resistance, as there is a linear relationship between the number of calls to first contact and background variables and a number of survey variables. This linear relationship disappears at the end of the continuum for the small number of cases that required a large number of contact attempts, possibly because of prolonged spells away from home. Section 7.3.3 shows that with a higher number of calls and higher contact rates, the response group becomes more

similar to the net sample of the Dutch population or includes more persons who are generally considered to be hard to reach. This should increase the precision of the survey outcomes and, in the case of nonresponse bias, also the accuracy (see section 2.3).

In summary, this chapter has shown that high contact rates can be achieved, how they can be achieved and that this seems to improve survey quality. Of course, once contacted, sample units have to cooperate. Late contacts do not seem to be less cooperative than early contacts. More will be said about cooperation in the next chapter.

8 Reluctance to cooperate

8.1 *Measuring reluctance: beyond the number of contacts*

Once contact has been established the sample person or household will be asked to cooperate in the survey. The process of obtaining cooperation and the analysis of the willingness - or reluctance - to cooperate differs from the process of establishing contact and the analysis of contactability in several ways, as explained out in section 2.6 and figures 2.3 and 2.5. The contactability of sample units can be influenced within certain limits by interviewers as they may base their call patterns on local knowledge or information obtained at earlier calls. Each call basically has two outcomes, contact or noncontact, and apart from the timing, subsequent calls on a household can be considered as independent events. It was shown in the preceding chapter that analysing contactability is far from simple. The process of obtaining cooperation and the analysis of reluctance is even more complicated, as each contact with a household can result in an interview (or at least an appointment), a final refusal or a next try. The choice between the latter two outcomes depends on the interviewer or the fieldwork organization. Refusals that seem soft and promise future cooperation are generally re-issued and contacted again (and again), whereas adamant, hard refusals are left alone. The model in figure 2.5 shows why it is difficult to interpret the number of subsequent contacts in a survey as an indication of the reluctance of sample persons to participate in a survey as it is a function of the outspokenness of the first refusal(s), the decision of the survey organization to re-approach the refuser, and the cooperation at later contacts. Formulated differently, a survey will comprise few reluctant respondents 1) if first contacts are very successful (and many result in interviews) or 2) if first contacts are very unsuccessful (and result in many not to be re-approached refusals), or 3) if field efforts are limited and no second contacts are made. As the number of contacts to final cooperation or refusal is not a straightforward measure of reluctance, other indicators of reluctance will be used in this chapter.

Another difference between ease of contact and reluctance to cooperate is the involvement of the prospective respondent. In face-to-face surveys, sample households will often be unaware of interviewer calls when nobody was at home. Many unsuccessful calls will therefore neither cause irritation nor admiration for interviewer perseverance. Each subsequent contact with (initially) reluctant sample persons, however, can be seen as the prolongation of a previous contact and the content of the interaction between sample person and interviewer will depend on the content of earlier contacts. In some cases, it will be better to delay the request for cooperation until a subsequent contact, which is what experienced interviewers do when they withdraw if they feel that they are calling at an awkward moment, and ask for cooperation at a second contact. Less experienced interviewers may not stop in time and will receive a refusal

at the first contact (see for instance Groves and Couper, 1998, pp. 38-39 and section 4.3.5). Experienced interviewers may therefore produce more contacts with sample persons and less experienced fewer contacts and more refusals.

Finally, the analysis of contactability and reluctance differs because the degree of contactability of (almost) all sample units can be determined since the final noncontact rate in face-to-face surveys can be kept down to 3% (see chapter 10). Reluctance to cooperate, on the other hand, is characterized by a dichotomy between more or less cooperative respondents and final refusals whose rate of reluctance will be difficult to assess.

In this chapter, the reluctance of respondents is investigated in several ways. Firstly, section 8.2 describes the process of obtaining cooperation outlined in section 2.6.3, explores the outcomes of the first and subsequent contacts with the sample household, assesses the likelihood of cooperation at subsequent contacts according to the outcome of the previous contacts, interviewer evaluations of future success and the deployment of new interviewers, and compares field efforts between regions. Section 8.3 tries to identify which respondent characteristics are related to reluctance indicated by a dichotomy between immediately cooperating and other households. A continuum of reluctance model is tested by combining several indicators of reluctance, namely the outcome of the first contact and the interviewer's assessment of future success after a temporary refusal. This section also presents a classic 'classes of participants' model in which contactability and reluctance to cooperate are combined and evaluates whether refusal conversion leads to more accurate survey outcomes. Section 8.4 presents information on the quality of survey outcomes obtained from amenable and reluctant respondents, and section 8.5 summarizes the results and answers the questions from section 5.5.

In section 2.6.3 a distinction was made between persons who were not able to cooperate and persons who refused to cooperate. Chapter 6 argued that noncooperation as outcome of a contact could be a 'not able' for different reasons, and a downright refusal. The reasons for refusal have not been coded. 'Not able' was subdivided into four categories, namely because of being ill, being busy, language problems and other reasons. Being ill and language problems are standard classes of not being able, where language problems will be a permanent condition and being ill either a transitory or a permanent condition. Noncooperation because of being busy is more often considered to be a situational reason for refusal than a real 'not able' (see section 4.2.2). In this study the target population excluded persons with language problems (see section 6.2.2) except for a very small number of households that could be interviewed anyway. No distinction could be made between those who were clearly mentally or physically unable to cooperate and those who were temporarily unwell. This is complicated by the fact that one person only was interviewed face-to-face and that drop-off questionnaires could be filled in by a responsible adult. In this

chapter the category 'unable' should mainly be interpreted as a situational reason for refusal, although a number of persons who were too ill to cooperate at the first contact will really be unable to be interviewed.

8.2 *Contact outcomes and history*

8.2.1 *Outcomes of subsequent contacts*

Table .1 Outcomes of first contact and final outcome for all contacted sample households

	outcome first contact	final outcome
interview	37.0	66.1
appointment	14.3	
broken off/incomplete	5.3	5.7
not able, ill	1.7	0.6
not able, busy	2.2	0.3
not able, language problems	0.1	0.1
not able, other reasons	3.8	2.0
refusal	35.5	25.2
N (=100%)	9261	9261

Figure 2.4.1 gave an overview of the possible outcomes of interviewer-household interactions. Table 8.1 presents the outcome of the first contact between interviewer and household and the final outcome of the AVO1999 survey, and shows that 37% of the first contacts resulted in an interview, 14% in an appointment, 5% in a broken-off interview, 36% in a refusal and 8% in 'not able' for other reasons. This means that at the first contact 56% of the households cooperated to a greater or lesser extent (interviews, appointments and broken-off interviews). This does not imply that the final response rate would have been 56% based on this initial success, as initial cooperation does not guarantee a complete response set.

Table 8.2 shows that from those households who immediately grant an interview (first row), 4% ends up with an incomplete set, which means that not all household members filled in a drop-off questionnaire. The appointment resulted in complete cooperation only in 82% as 6% of these households delivered an incomplete set and 12% refused after all, or said they were not able. Of the broken-off interviews only one third were successfully resumed, 48% remained incomplete and 19% were met by a later refusal. Added together, this means that the cooperation rate would have been (0.96 x 37% + 0.82 x 14.3% + 0.33 x 5.3%) = 49%, rather than the 66% in the second column of table 8.1, if only those households remained in the sample that appeared to be cooperative at the first contact. Response rates can thus be substantially enhanced, and refusal rates reduced, by re-contacting initially reluctant sample households.

Table .2 Outcome first contact (row) and final outcome for all contacted sample households (column) (%)

	interview	not able				refusal	incom-plete	total
		ill	busy	language	other			
interview	96						4	100
appointment	82	0	0	0	1	11	6	100
broken-off	33			0	0	19	48	100
not able, ill	48	25		1	2	22	1	100
not able, busy	56	0	8		1	32	2	100
not able, language problems	40			20		30	10	100
not able, other reasons	47	1	0	0	28	22	2	100
refusal	37	0	0	0	2	72	3	100
total	66	1	0	0	2	25	6	100

In a small number of cases language problems made a sample unit ineligible (see table 6.3). In an even smaller number of cases (10 households, 0.11% of the eligible sample) persons who first appeared to have to be excluded from being interviewed were considered to be eligible after all, and 4 of them were successfully interviewed. Other reasons for initially not being able to be interviewed in table 8.1 are illness (1.7%), being engaged in activities (2.2%) and being otherwise engaged (3.8%). These reasons will be considered as situational factors for (temporary) refusal in the remainder of this chapter (see section 4.2.2). If a household refused because of these situational factors, the chances were around 50% according to table 8.2 that it would cooperate eventually. Those who were ill at the first contact remained so (or were ill again) in one quarter of the cases, and refused in almost another quarter. Those who were to busy to cooperate were finally interviewed in 56%, remained busy in only 8% and refused in one third of the cases. Those who were otherwise engaged were interviewed in almost half of the cases (47%), remained otherwise engaged in 28% and refused in 22%. The table shows that the subsequent cooperation rate is higher for situational factors (ill, busy) than for a downright refusal that may be more related to the type or topic of the survey or the type of sponsor. Those who refused at the first contact (35.5%) were converted only in 37% and remained refusals (either because they refused again or because there was no subsequent contact) in 72%.

Table 8.2 presented the final outcome for each sample household dependent on the outcome of the first contact. Figure 8.1 pays attention to intermediate stages and gives the outcomes of each subsequent contact dependent on the outcome of the first contact, except for those cases when an interview had been conducted at the first contact. Figure 8.1a shows what happened if the first contact resulted in an (initial) refusal. A second contact was established with about two thirds of the refusing

households. The others were either left alone, or never found at home and thus became final refusals. At the second contact, one third of the (temporary) refusals agreed to be interviewed, one third refused again or gave another reason for non-compliance and were recorded as final refusers, and one third remained in the field and were contacted a third time. At the third contact, 43% of the two-time refusers agreed to be interviewed, 40% dropped out definitively and 15% were to be contacted a fourth time. Of these (126) households, 59% agreed to be interviewed, and 16% stayed in the field. At the fifth contact, 12 of the 20 rather persistent refusers cooperated, seven were final refusals and one was interviewed at the sixth contact.

Figure 8.1.a Outcome first contact refusal

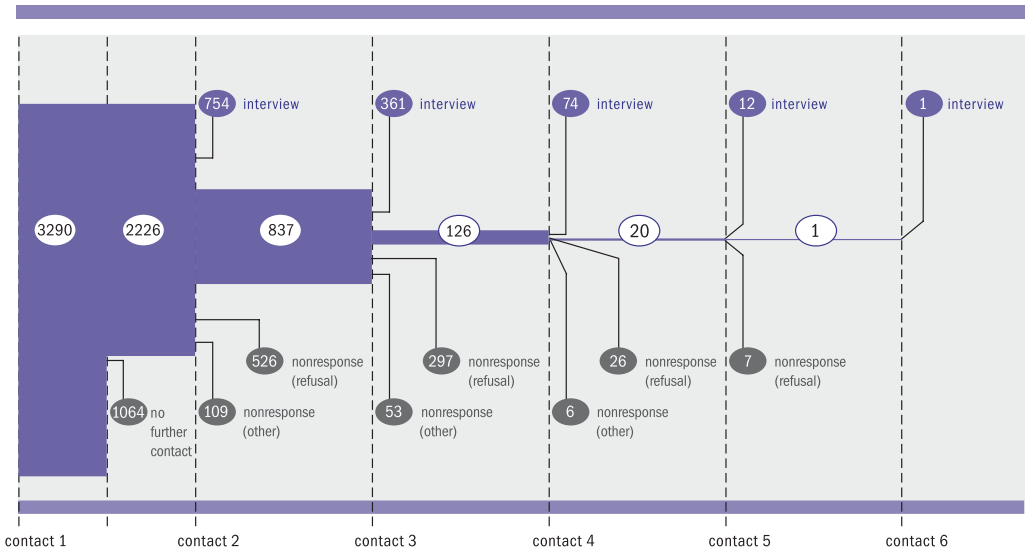


Figure 8.1a shows quite clearly that after frequent contacts the cooperation rate is high. This should not be seen as an indication that repeatedly approaching sample persons makes them more pliable, but rather that interviewers only go back to those households that seem promising. Figures 8.1b, c, d and e present in a similar way the field history of those households with whom an appointment was made at the first contact or where the first contact was with someone who was too ill, too busy or too otherwise engaged to be interviewed. At each subsequent contact a substantial part of the earlier refusers was converted. The figures show that initial reasons for refusal should not always be taken at face value. Of the 205 householders who indicated at the first contact that they were too busy, 56% were interviewed at a subsequent call, 8% received no further calls or remained too busy, and 37% refused to participate at a subsequent call. This indicates that 'not able, too busy' may be a polite way of refusing cooperation. Busy respondents were followed up much more than ill ones. This may be due to the fact that illness sometimes may be a more permanent state of life (think about mental and physical incapacity) than a transitory condition as being engaged.

Figure 8.1.b Outcome first contact appointment

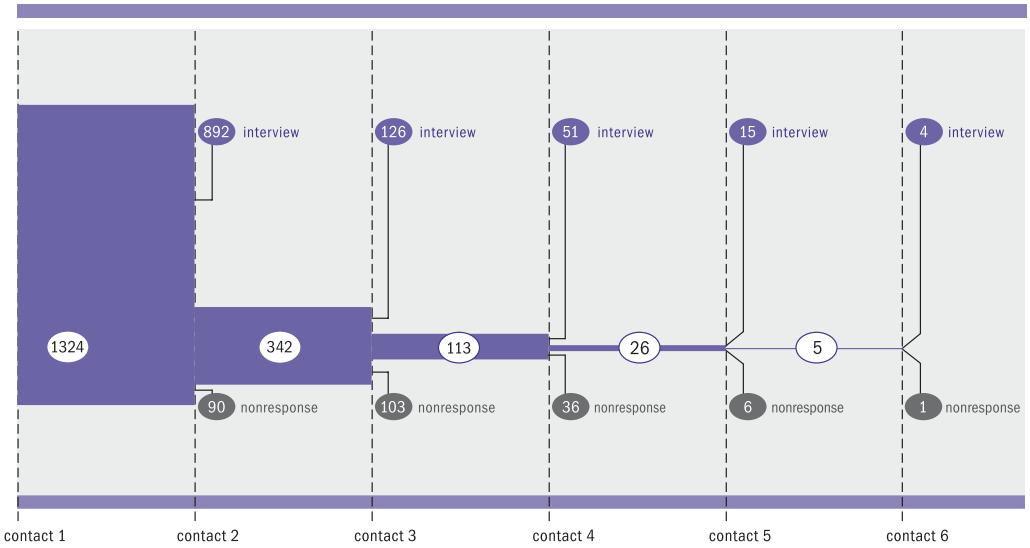


Figure 8.1.c Outcome first contact not able, ill

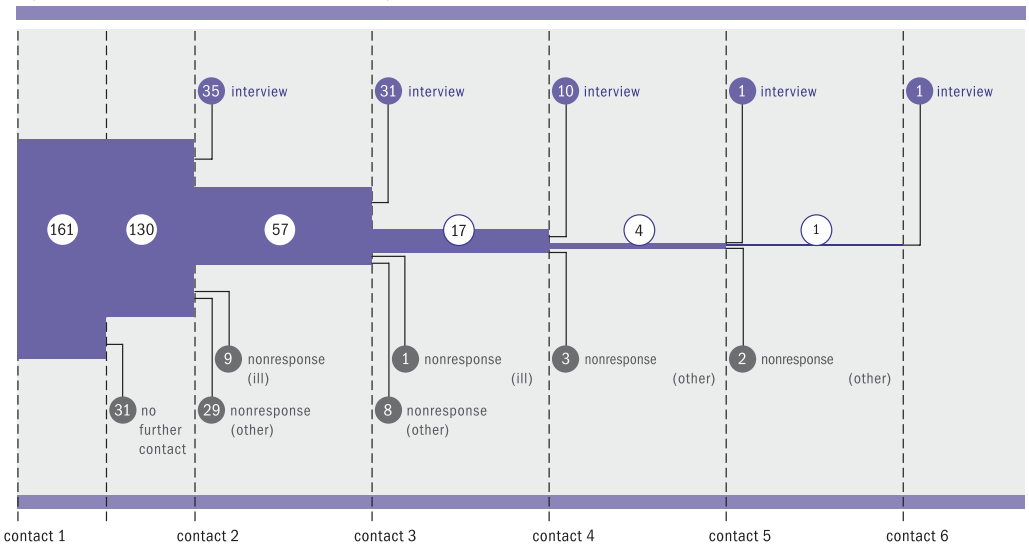


Figure 8.1.d Outcome first contact not able, busy

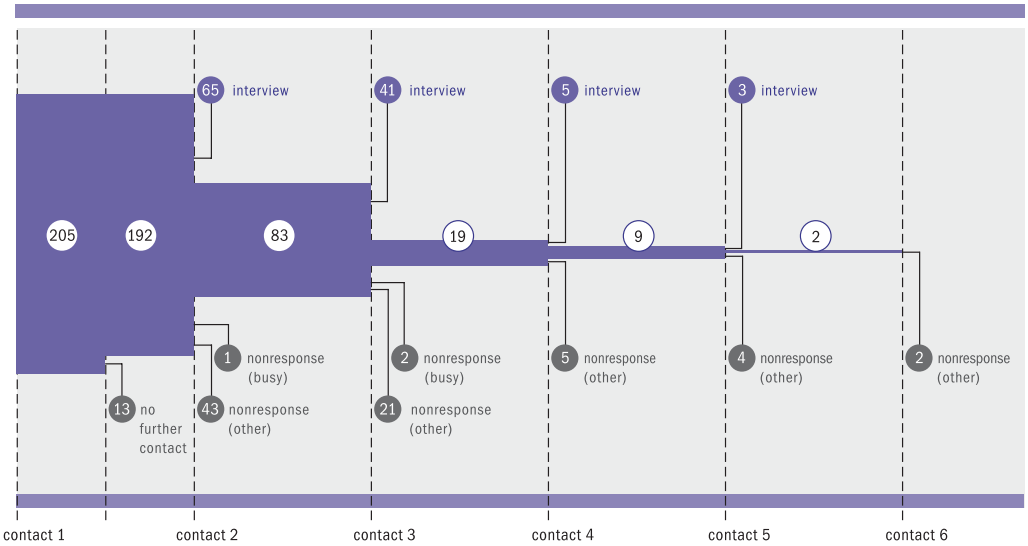


Figure 8.1.e Outcome first contact not able, other reasons

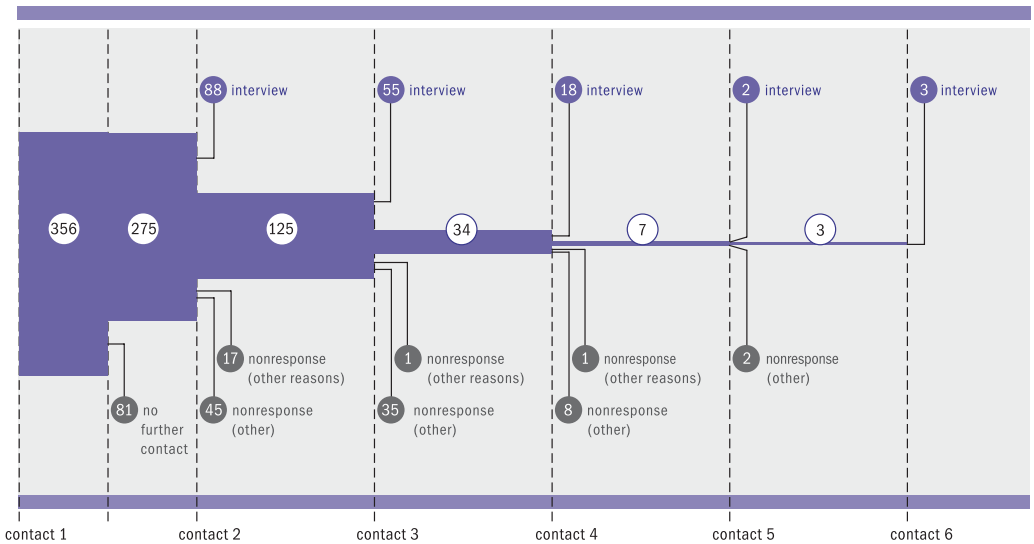


Table 8.3 presents the multiple reasons for nonresponse given by 13 extremely reluctant households that were contacted six or seven times. These persistently reluctant households were not particularly hard to contact (otherwise the number of contacts could not have been this large). The table suggests that re-visiting noncompliant households is worthwhile as most of these finally granted an interview. What should kept in mind, again, is that these results do not prove that an infinite number of contacts will finally result in an interview, but that to achieve the target response rate interviewers kept returning to households which they felt would ultimately be converted, and presumably with success. When they had little trust in a happy ending, the sample address may not have been re-issued, and then of course no interview was conducted.

Table .3 all records for households with 6 or more contacts[a]

call	con-tact	call record
6	6	not able, busy (thrice), interview, broken off, not willing (refusal)
6	6	appointment (thrice), interview, not willing (refusal), interview
7	6	not able (other), not willing (other), not able, (ill) (twice), not at home, not willing (refusal), interview
7	6	not able (other), appointment, broken off, not willing (refusal), not at home, appointment, interview
7	6	not able (busy), not willing (other), not willing (refusal) (thrice), not at home, not willing (refusal)
8	6	appointment, not willing (other), not able (other), not willing (refusal), not at home (twice), appointment, interview
8	6	broken off, not willing (other), not willing (refusal) (twice), not at home (twice), appointment, interview
8	6	appointment, not at home, appointment, broken off, not willing (refusal), not at home, appointment, interview
9	6	appointment (twice), not willing (other), not at home, not willing (refusal), not at home (twice), appointment, interview
9	6	not at home, not able (other), not willing (other), not able (language), not willing (refusal), not at home (twice), appointment, interview
9	6	not at home, appointment, not willing (other), not at home, appointment, not willing (other) (twice), not at home, not willing (other)
9	7	not able (ill) (thrice), not willing (other), not willing (refusal), not at home (twice), appointment, interview
11	6	not at home (thrice), not willing (other), not able (other) (twice), not willing (refusal), not at home (twice), appointment, interview

a Not willing (refusal) indicated a general reason for nonresponse, whereas under 'not willing (other)' the interviewer could record specific reasons. As the recorded information has not been keyed, the distinction between these two reasons for refusal provides no additional information and will be ignored.
In several cases the interviewer recorded more than one interview. The reason for this is not clear, but this most likely concerns broken-off interviews.

As explained in section 6.2.2 the fieldwork for the AVO took place in three subsequent waves. After each wave all eligible addresses that had not resulted in an interview were appraised by the fieldwork organization to decide whether or not

they should be re-issued. The decision to re-issue an interview after fieldwork wave 1 and wave 2 was based on the reason for refusal, the appraisal of interviewer quality, both in general and at that particular moment, and the interviewer's assessment of the possible success of a future contact. Table 8.4 presents the outcomes of contacts according to the judgment of the interviewers after excluding those for whom no assessment had been made or should have been made because an interview was conducted within the wave. Of those households for which an interviewer indicated at least once that an interview would be absolutely impossible, one third was left alone, one third did not cooperate at a subsequent contact and one third resulted in a successful interview after all. The success rate is slightly higher (39%) for those households for which the interviewer recorded at least once that they would rather not be interviewed. In this case, more interviews were re-issued. The situation is even better for those about whom the interviewer thought once or twice that they would possibly or probably cooperate at a later stage (51% interviews). The least successful were the results for those about whom the interviewers recorded that they would absolutely not cooperate after they had recorded a different assessment at the end of an earlier wave. However, 23% of this group still ultimately participated. From these results one cannot determine how successful interviewers were in assessing the probability of future success, as these assessments partly determine whether or not the noncompliant households were contacted again or left alone. These outcomes might partly be considered a self-fulfilling prophecy but still indicate that contacting temporary refusals can substantially increase response rates.

Table .4 ssessment of future success and final outcome[a]

	not re-issued	re-issued, interview	re-issued, no interview	total
N				
absolutely not	602	612	592	1806
absolutely not (after other assessment in earlier wave)		55	180	235
rather not	190	295	265	750
possibly, probably later (wave 1,2)	26	107	76	209
total	818	1069	1113	3000
%				
absolutely not	33	34	33	100
absolutely not (after other assessment in earlier wave)		23	77	100
rather not	25	39	35	100
possibly, probably later (wave 1,2)	12	51	36	100
total	27	36	37	100

a Excluding 6213 households for which no success assessment was made (of which 3182 interviews at first contact) and 50 households with a mixed set of assessment values across the three waves

After an initial refusal, the preferred strategy seems to be to re-issue the interview to a new interviewer (see section 2.6.3). Table 8.5 shows the results of subsequent contacts after an unsuccessful earlier contact in the AVO1999. This table distinguishes

between subsequent contacts by the same interviewer and by a different interviewer.[1] After a 'not able' in a previous contact the same interviewer called again in more than two-thirds of the cases (479 out of 677). Both original (35%) and new (33%) interviewers achieved an interview in approximately one third of these cases. After a previous refusal, the situation is quite different. Now, the original interviewer was used in fewer than one third of cases (853 out of 2,780). This seems illogical, as the original interviewer had a much higher success rate (52%) than the new one (28%). The simplest explanation for the relative success of the original interviewer is that the fieldwork organization might be more inclined to send back the original interviewer in the easier cases where the refusal is not considered to be adamant, and to employ a new interviewer in the more difficult cases where they did not expect the original interviewer to succeed. Here, too, operational procedures aimed at response enhancement and cost reduction may limit the possibilities of nonresponse analysis that would benefit from a random choice in re-issuing surveys between the original and a new interviewer.

Table .5 Outcome contacts for previous contact 'not able' and 'not willing' for identical and different interviewers across subsequent contacts[a]

	interview	appointment	broken-off	not able	not willing	total	N contacts
previous contact with the same interviewer							
not able	35	11	3	18	33	100	479
not willing	52	11	2	3	32	100	853
total	46	11	2	8	32	100	1332
previous contact with the different interviewer							
not able	33	11	2	13	40	100	198
not willing	28	10	1	5	56	100	1927
total	29	10	1	6	54	100	2125
previous contact with the same or different interviewer (total)							
not able	35	11	3	16	35	100	677
not willing	36	10	1	4	48	100	2780
total	36	10	2	7	46	100	3457
total including contacts for which no interviewer identification was available							
not able	35	12	3	13	37	100	765
not willing	37	10	1	4	47	100	3262
total	36	10	2	6	45	100	4027

a For all contacts that were preceded by an a previous contact

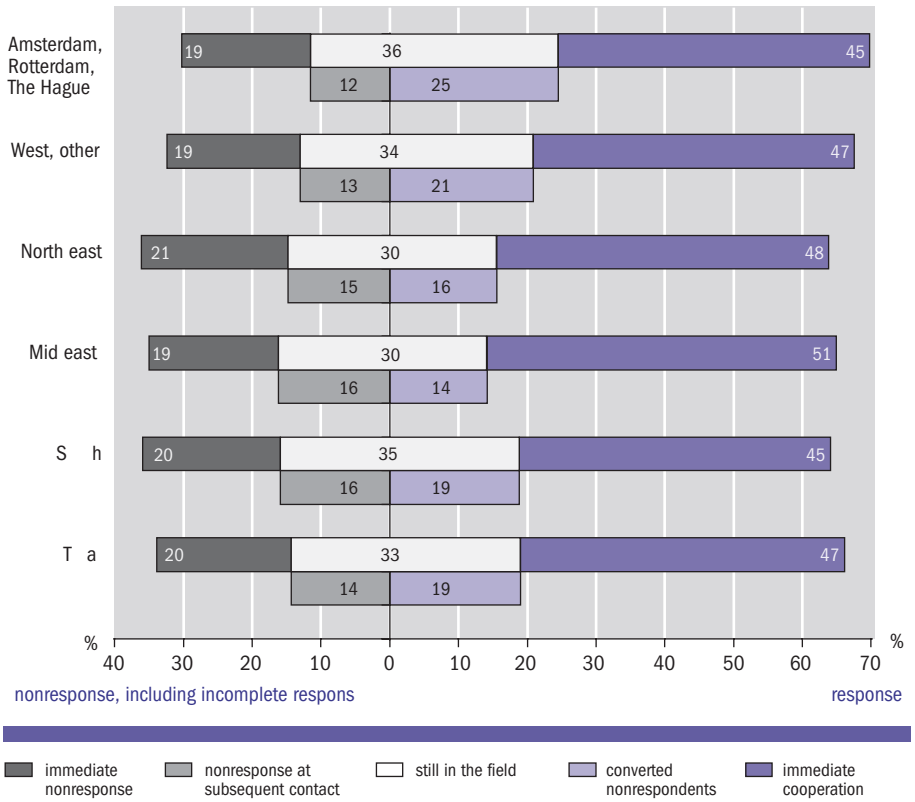
8.2.3 *Regional differences in obtaining cooperation*

Urbanicity is generally considered to be a major correlate of survey nonresponse, both of contactability and cooperation (see section 3.2 and 3.3.4). Figure 7.2 in

1 These results are not complete, as in particular regions no interviewer identification was available (see section 6.2.2)

section 7.2.1 showed that interviewers in big cities anticipated that urban residents would be at home less during the day. Figure 8.2 presents other regional differences in field efforts and outcomes. The lower bars show that at the first contact with approximately 47% of the sample households a successful interview was completed or an appointment was made that resulted in a successful interview. In some regions the initial cooperation was lower (big cities and southern part of the Netherlands, 45%), and in other parts of the country somewhat higher (mid-eastern region: 51%). Approximately 20% of the households were written off as final refusals after one contact, either because they were never called on again or because they were called on but not found at home. One third (33%) of the households remained in the field after the first contact and were re-contacted later. This varied from 30% in the eastern part of the Netherlands to 36% in the big cities. Of these 33%, 19% (=57%) granted in interview.

Figure 8.2 Fieldwork efforts and region



What is most striking in figure 8.2 is the success of refusal conversion in Amsterdam, Rotterdam and The Hague. Of the 36% households still in the field after the first contact 12% refused again and 25% (two-thirds of the re-contacted households)

Table .6 Fieldwork outcomes per region

	% evening calls (first call)	# calls to first contact	% cooperation immediate	% cooperation final	% incomplete	# calls co-operative	# calls converted	# calls 2+ refusals	# calls 1 refusal	# contacts co-operative	# contacts converted	# contacts 2+ refusals	# contacts 1 refusal
Amsterdam, Rotterdam, The Hague	18.5	2.76	45	70	3.3	3.59	4.95	6.00	4.52	1.26	2.53	2.47	1.33
west, other	6.9	2.16	47	68	4.6	2.71	4.64	5.06	3.77	1.34	2.61	2.42	1.32
north east	6.8	1.91	48	64	6.2	2.32	4.09	4.66	3.08	1.34	2.52	2.64	1.36
mid east	16.0	1.81	51	65	6.9	2.28	4.13	5.22	3.20	1.39	2.70	2.69	1.47
south	17.4	2.02	45	64	5.4	2.42	4.47	5.40	3.30	1.30	2.66	2.53	1.30
mean	13.1	2.12	47	66	5.2	2.63	4.53	5.26	3.57	1.33	2.61	2.54	1.35
difference big cities and mean	5.4	0.64	-2	4	-1.9	0.96	0.42	0.74	0.95	-0.07	-0.08	-0.07	-0.02

granted an interview, whereas in the rural mid-eastern region (Flevoland, Overijssel, Gelderland) the figure was less than half (14% of the 30% re-contacted households). These differences could be caused either by respondents in the less urbanized areas being less compliant than big city-dwellers, or by interviewers in big cities being more persistent than in the less urbanized areas.

As shown in section 7.2.1, interviewers in the big cities more often started with an evening call. Nevertheless, as table 8.6 shows, they had to make more calls to get in touch with the sample households (2.76 calls compared to 2.12 calls on average). According to figure 8.2 the percentage of households that cooperated without once refusing or saying they were not able was smaller in big cities (45%), but still the final response rate here is the highest of all regions. This is presumably brought about by making many calls at the sample households.

The bottom row of table 8.6 shows that, compared to the national average, immediately cooperative respondents in big cities received one (0.96) call more than average, converted refusers 0.42 call more, refusers who were re-contacted and refused again 0.74 call more, and refusers who had not been contacted again 0.95 call more. The number of contacts is remarkably similar across the regions: about 1.33 for immediately cooperative and one-contact refusers (it would have been 1 if appointments and broken-off interviews had not been taken into account), and approximately 2.6 for converted refusals and those who refused at least twice.

The more rural north-eastern area of the Netherlands shows precisely the opposite pattern from the urban areas in the western region. Households were fairly easy to reach, despite the lower rate of evening calls. Immediate cooperation was fairly high, the final cooperation rate rather low (64%), possibly due to the relatively low number of calls on the less compliant respondents. In addition, in big cities the 19% of refusals with whom there was no further contact can be subdivided into 7% who received no further calls and 12% who were not found at home after a refusal. In the north-eastern provinces, by contrast, the refusals without further contact received no further calls in 14% of cases and were not found at home in 7%. Finally, table 8.6 shows that the percentage of incomplete sets (a complete set comprises the face-to face household questionnaire and self-completion questionnaires for all members of the household aged 6 years and older) was lowest in the big cities and highest in the eastern parts of the country.

In summary, the results seem to indicate that AVO interviewers in big cities overcompensated for low accessibility and low amenability by trying harder. As a consequence, reluctant respondents – those who had to be converted from an earlier refusal – are overrepresented in big cities, both because there were so many refusals to begin with and because interviewers made great efforts to achieve high response rates.

8.3 Who is reluctant?

In the introduction to this chapter it was mentioned that the number of contacts is a poor indicator of reluctance, as re-contacting sample households after a previous refusal is a strategy for obtaining the cooperation of promising cases, whereas adamant refusals are more likely to be left alone. The overview of regional differences given above confirms that a larger number of contacts may be due both to a high rate of initial reluctance and to a high rate of refusal conversion. Therefore, the number of contacts cannot reflect a continuum of resistance. Section 8.3.1 explores whether there are better measures of reluctance, and section 8.3.2 presents a classic 'classes of participants' model in which accessibility and amenability are combined. Those background and survey variables will be included in the models which could be expected to be related to reluctance according to the literature presented in sections 3.2 and 4.2, which were part of the survey and which showed a relationship with reluctance. Section 8.3.3 reviews the analyses and discusses whether refusal conversion did improve data quality in the AVO1999.

The data in this chapter refer to respondents only. An analysis of the reluctance of all sample units, including final refusals, showed no relationship with any of the variables that are available for both respondents and nonrespondents. Urbanicity, neighbourhood and dwelling characteristics all showed desperately insignificant correlations with 'cooperation'.[2] This may be due to the fact that reluctance is poorly measured by this variable, because reluctance is related to household or individual characteristics rather than to regional factors (see section 3.4.2 on urbanicity) or that field efforts were unevenly distributed across regions (see previous section). In the next chapter final refusers are brought back into the analysis of the follow-up survey among refusers.

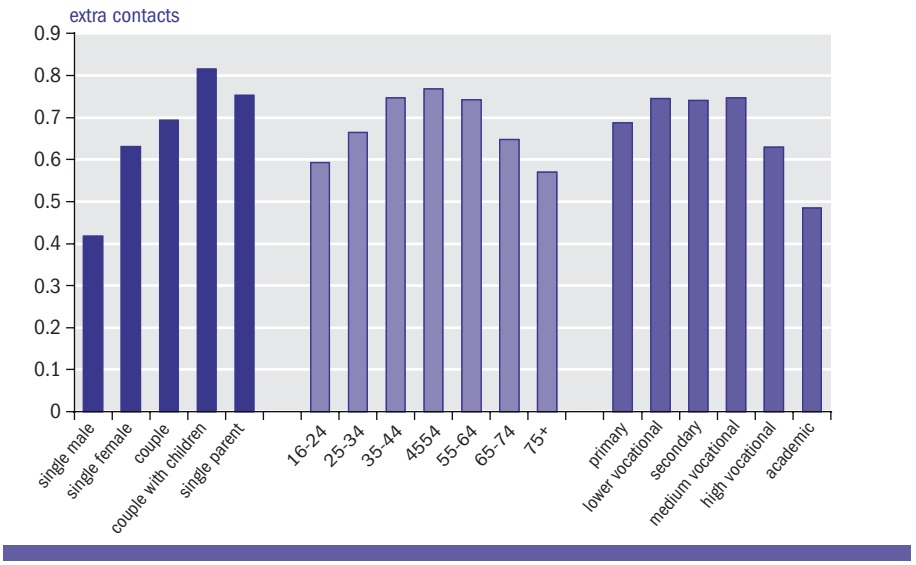
8.3.1 Degrees and types of reluctance

In section 3.3 a number of background variables were presented that are often considered to be related to reluctance to cooperate in surveys. As an example, figure 8.3 represents the number of contacts in addition to the one at which the interview was conducted for three variables that showed a relationship with reluctance in the AVO1999, namely family composition (including the sex of single persons), the age of the responsible adult with whom the interview was conducted, and his or her level of education.[3]

2 Cooperation had the following categories: immediate cooperation, cooperation after previous refusal/not able, refusal without subsequent contact (possibly including unsuccessful contact attempts after refusal), refusal at first and subsequent contact(s))

3 As the AVO is a household survey, sex is only given for single-person households. In families, any knowledgeable adult could be the main interviewee, and another adult could have refused earlier. In single person families, however, no such confusion is possible.

Figure 8.3 Number of contacts (in addition to interview) according to family composition, age and education



It shows that single males required 0.4 contacts in addition to the one contact required to conduct an interview, and that couples with children 0.8 extra contacts. Age shows a curvilinear relationship with fewer additional contacts for both younger and older respondents, and more highly educated respondents also receive fewer subsequent contacts. Family composition and education will come back in the models of reluctance presented in this chapter. Age had no independent effect and will play no further role in the analysis of reluctance.

The logistic regression in table 8.7 (see annex 7.1) predicts initial reluctance, defined as having at some time refused to cooperate, from socio-structural characteristics and survey variables. Only independent variables that had a significant relationship with reluctance have been included in the model. The results indicate that the temporary refusals include relatively few males, single persons, persons with a higher education level, higher-income households and ethnic minorities. More temporary refusals than expected live in big cities, which is in line with the results of extended interviewer efforts in big cities mentioned in the previous section, and fewer live in well-maintained dwellings and neighbourhoods and detached dwellings. Among the reluctant respondents there is a relatively large number of who participate in classical cultural activities. There is a small negative effect for contactability: households that showed some reluctance had received slightly fewer calls to first contact. Instead of concluding that there is an inverse relationship between accessibility and amenability, it should be mentioned that this is most likely a consequence of thefieldwork logistics discussed in section 7.4. If it takes many calls to contact a respondent for the first time, there simply

is not much time left for refusal conversion strategies, as these require additional calls in the later stages of the fieldwork.

Table . aving ever refused, logistic regression among respondents[a]

	β	exp β	S.E.	Wald	Sig.
number of call first contact	–0.13	0.88	0.02	36	0.000
Amsterdam, Rotterdam, The Hague	0.54	1.72	0.08	43	0.000
good maintenance	–0.18	0.84	0.07	6	0.017
detached dwelling	–0.37	0.69	0.10	14	0.000
male	–0.78	0.46	0.07	126	0.000
single	–0.63	0.53	0.08	65	0.000
ethnic minority group[b]	–0.81	0.44	0.24	11	0.001
higher education	–0.45	0.64	0.10	22	0.000
higher income	–0.26	0.77	0.08	9	0.002
cultural participation (df=3)				14	0.003
both classical and popular	–0.01	0.99	0.08	0	0.902
classical	0.23	1.26	0.08	8	0.006
popular	–0.10	0.90	0.10	1	0.299
none (ref.)					
constant	–0.39	0.67	0.08	25	0.000

a Chi^2 369 (df=12), Cox & Snell R^2 0.058 Nagelkerke R^2 0.088 N=6,125
Reference group cultural participation: no participation; β's reference categories: 0

b Country of birth respondent, respondent's mother or respondent's father Morocco, Turkey, Surinam, Dutch Antilles or Aruba

Table 8.8 presents a multinomial regression that distinguishes between five different levels of reluctance at the first contact. The most cooperative group (and reference group, see annex 7.1) are those who cooperated immediately. Slightly more resistance was displayed be respondents for whom an appointment or a broken-off interview preceded the interview. Even more reluctant were those who said at the first contact that they were not able to cooperate. This can be seen as a situational refusal (section 4.2.2). The most difficult group were those who actually refused at the first contact. The latter are subdivided into the highest level of reluctance about whom the interviewer recorded at the end of the first wave that there was absolutely no chance that they would ever cooperate ('hard' temporary refusers), and the rest (now called 'soft' temporary refusals). These categories are assumed to reflect increasing reluctance. This assumption can be tested by finding out if there is a linear relationship between these levels of reluctance and those background and survey variables that are generally considered to be related to survey cooperation (see section 5.2). Only if this is true, it is likely that the final refusers are similar to the most persistent temporary refusals. Otherwise, either these categories do not reflect an increasing degree of resistance, or they might indicate different classes of participants, or reluctance of respondents is too confounded with field strategies. The categories in the model are based on the outcome of the first contact. Other outcomes of intermittent contacts may have preceded the final interview.

The upper part of table 8.8 shows the percentage of households that belong to a particular reluctance group according to the background and survey variables. For instance, 16% of those who immediately granted an interview were big-city dwellers, 42% were males, 20% had a higher education, etc. These percentages can be compared with the share of these groups in other reluctance categories and their share among all respondents. A first comparison shows, for instance, that big-city dwellers are overrepresented among soft temporary refusals and underrepresented among hard refusals (23% and 14%, respectively, compared with 16% in total), and that men and the better educated are underrepresented among both soft and hard refusals.

The lower part of table 8.8 shows the results of the multinomial regression itself. The negative effect of the number of calls to first contact for all initially reluctant groups shows that they received fewer calls to make first contact than immediately cooperative respondents, as was also shown in table 8.7. Those respondents with whom the outcome of the first contact was an appointment or a broken-off interview hardly differ from those who immediately grant an interview. They include fewer males and fewer singles, groups that are underrepresented in each initially reluctant group. Surprisingly, participation in popular culture is significantly higher in this group, although the actual differences are very small.

The respondents who were not able to participate at the first contact because of illness, being too busy or for other reasons, can hardly be distinguished from the reference groups in terms of the variables in the model. Being not able might be a situational reason for refusal that has probably less to do with a dislike of surveys, the sponsor or the topic (see section 4.2.2) and therefore the higher 'resistance' of this group might just be coincidental and be due to nothing more than bad timing. They do live slightly more often in big cities and less often in well-maintained neighbourhoods and dwellings, and are less often male.

The soft and hard temporary refusals show the largest differences compared with the immediately cooperative respondents. They less often live in well-maintained neighbourhoods and dwellings and in detached homes, are less often male, single or members of ethnic minority groups, less often have a higher education and a higher income and more often take part in classical cultural activities. This is all similar to the distinction never/ever refused in table 8.7. The only distinction between soft and hard temporary refusals is that the former were overrepresented in big cities and the latter not. It is not clear whether this has something to do with respondent cooperation or with interviewers in big cities being more likely to classify a refusal as willing to cooperate at a subsequent try, or with big-city interviewers trying more often to convert soft refusals (and succeeding).

Table . Gravity of temporary refusal: multinomial regression among respondents (n 6125)[a]

	Immediate interview (ref.)	appointment/ broken off		not able (situational refusal)		soft temporary refusal		hard temporary refusal		total
	n									
	3301	1250		362		756		456		6125
	mean									
number of call first contact	2.39	1.75		1.83		2.08		1.94		2.16
	%									
Amsterdam, Rotterdam, The Hague	16	13		18		23		14		16
good maintenance	27	28		22		23		22		26
detached dwelling	14	17		12		10		11		14
male	42	37		34		23		27		37
single	31	24		28		20		21		27
ethnic minority group[b]	3	2		3		1		1		3
higher education	20	22		17		12		11		18
higher income	23	27		25		19		17		23
cultural participation both classical and popular	30	32		34		29		28		31
classical	21	21		23		26		27		22
popular	17	19		14		16		13		17
		β	Exp β	β	Exp β	β	Exp β	β	Exp β	
intercept		–0.365		–1,546		–0,732		–1,008		
number of call first contact		–0.301	0.740	–0.264	0.768	–0.140	0.869	–0.181	0.835	
Amsterdam, Rotterdam, The Hague		–0.020	0.980	0.265	1.304	0.613	1.845	0.041	1.042	
good maintenance		–0.104	0.901	–0.365	0.694	–0.196	0.822	–0.282	0.754	
detached dwelling		0.118	1.125	–0.123	0.884	–0.364	0.695	–0.368	0.692	
male		–0.253	0.776	–0.354	0.702	–0.847	0.429	–0.665	0.514	
single		–0.247	0.781	–0.084	0.920	–0.713	0.490	–0.604	0.547	
ethnic minority group		–0.042	0.959	0.081	1.084	–1.020	0.360	–0.833	0.435	
higher education		0.090	1.094	–0.222	0.801	–0.451	0.637	–0.479	0.619	
higher income		0.076	1.079	0.205	1.227	–0.257	0.774	–0.324	0.723	
cultural participation both classical and popular		0.232	1.261	0.231	1.260	0.142	1.152	0.090	1.094	
classical		0.103	1.109	0.156	1.168	0.352	1.421	0.278	1.321	
popular		0.295	1.343	–0.093	0.912	0.088	1.091	–0.202	0.817	

Table . (cont.) Gravity of temporary refusal: multinomial regression among respondents (n 6125)[a]

	sign.	Wald	sign.	Wald	sign.	Wald	sign.	Wald	X2
intercept	0.000	15.922	0.000	109.8721	0.000	48.98	0.000	63.39	
number of call first contact	0.000	119.15	0.000	35.8788	0.000	28.02	0.000	24.82	188
Amsterdam, Rotterdam, The Hague	0.843	0.04	0.081	3.0489	0.000	33.32	0.779	0.08	36
good maintenance	0.187	1.74	0.009	6.9115	0.050	3.84	0.023	5.18	13
detached dwelling	0.220	1.51	0.481	0.4972	0.009	6.89	0.027	4.90	16
male	0.000	13.04	0.003	9.0020	0.000	79.99	0.000	34.62	112
single	0.003	8.80	0.529	0.3966	0.000	46.93	0.000	22.85	66
ethnic minority group	0.847	0.04	0.806	0.0601	0.004	8.15	0.052	3.76	15
higher education	0.319	0.99	0.159	1.9848	0.000	12.18	0.004	8.53	26
higher income	0.384	0.76	0.160	1.9702	0.024	5.12	0.024	5.11	15
cultural participation both classical and popular	0.012	6.28	0.121	2.4041	0.211	1.56	0.514	0.43	33
classical	0.291	1.12	0.323	0.9774	0.002	9.67	0.039	4.27	
popular	0.004	8.28	0.611	0.2594	0.495	0.47	0.222	1.49	

a Contrast interview after appointment/broken off, not able, soft temporary refusal (not characterized as "absolutely no cooperation in future contact"), hard temporary refusal (characterized as "absolutely no cooperation in future contact"), reference category: immediate cooperation.
Chi-square 543 (df=48); Cox and Snell R^2 0.085 Nagelkerke R^2 0.092. reference categories 0 (opposite listed categories; no cultural participation).

b Country of birth respondent, respondent's mother or respondent's father Morocco, Turkey, Surinam, Dutch Antilles or Aruba.

Particularly interesting is the relationship between cultural participation and survey cooperation. Those who more often participate in popular culture (cinema, pop concerts, dance events) are overrepresented among those respondents who require an appointment or a second meeting to finish a broken-off interview. The converted refusals, however, show an overrepresentation of respondents who take part in classical cultural activities (concerts, theatre) only and not in popular culture. These effects remain after inclusion of background variables in the model that are related to cultural participation. One tentative explanation might be that those involved in popular culture are less available for interviews and have more busy schedules, and that participation in classical culture is related to being part of the cultural mainstream and thus being more receptive to interviewer persuasion to take part in a survey on government-sponsored services (among which classical cultural performances).

The largest effect in table 8.8, apart from contactability, is the effect of sex (male) and family composition (single). Males and singles are underrepresented among the more reluctant respondents, especially among those who refused at first contact, and even more so among the 'soft' refusals than among the 'hard' refusals of whom the interviewer assessment of future success was very negative. The negative effect of

the characteristics 'male' and 'single' suggests that these groups are more likely to cooperate immediately. Indeed they do, but that is not necessarily because they are more compliant. A better explanation will be given in section 8.3.3.

The evidence presented above does not seem to point to the existence of a continuum of reluctance. The number of contacts as an indicator of reluctance does not reflect the respondent's unwillingness and the levels of reluctance in table 8.8 do not have a linear relationship with background and survey variables. It could be assumed that the categories in table 8.8 reflect classes of reluctance rather than degrees of reluctance (see section 5.2). Section 8.3.3 reviews the assumptions and outcomes of the continuum of reluctance model after the presentation of a classes of participants model in section 8.3.2.

8.3.2 Classes of participants

This section follows the classic model in survey research in comparing classes of participants that reflect both accessibility and amenability (see section 5.2 and figure 5.1). A more complete classes of nonparticipants model, including final refusals, will be presented in the next chapter. In the present model in table 8.9 the underlying assumptions are that late respondents are similar to noncontacts (of which there are hardly any because of the high contact rate) and converted refusals similar to the unknown final refusers, and that late contacts, reluctant respondents and late and reluctant respondents differ in their own way from easy respondents.

Table .9 Net sample (respondents), dependent: response group, multinomial regression (n 6125)[a]

	no refusal 3+ calls	refusal 1-2 calls	refusal 3+ calls	no refusal 3+ calls	refusal 1-2 calls	refusal 3+ calls	total	
		β			exp β			
Intercept	–1.325	–0.627	–2.098	0.266	0.534	0.123		
listed telephone number	–0.279	–0.227	0.010	0.757	0.797	1.010		
Amsterdam, Rotterdam, The Hague	0.618	0.556	0.998	1.855	1.744	2.712		
detached dwelling	–0.295	–0.249	–0.920	0.745	0.780	0.399		
good maintenance	–0.260	–0.273	–0.236	0.771	0.761	0.790		
family composition[c]								
single male	0.579	–0.995	–1.105	1.784	0.370	0.331		
single female	0.453	–0.405	–0.282	1.572	0.667	0.754		
couple	0.264	–0.230	–0.037	1.302	0.794	0.964		
ethnic origin[b]	0.211	–0.859	–0.800	1.235	0.424	0.449		
student, employed	0.472	–0.188	0.012	1.603	0.829	1.012		
higher education	0.006	–0.412	–0.875	1.006	0.662	0.417		
cultural participationc								
both classical and popular	0.248	0.087	0.260	1.281	1.091	1.296		
classical only	0.193	0.236	0.463	1.213	1.266	1.588		
popular only	0.262	–0.059	0.298	1.300	0.942	1.347		
		significance			Wald		X2	df
intercept	0.000	0.000	0.000	145.49	31.28	142.62		
listed telephone number	0.001	0.012	0,939	11.97	6.35	0.01	15.4	3
Amsterdam, Rotterdam, The Hague	0.000	0.000	0.000	49.41	30.49	58.57	90.9	3
detached dwelling	0.006	0.025	0.000	7.52	5.05	17.81	28.2	3
good maintenance	0.001	0.001	0.068	10.55	10.17	3.32	18.0	3
family compositionc							145.8	9
single male	0.000	0.000	0.000	29.01	37.14	17.43		
single female	0.000	0.000	0.079	21.31	13.75	3.09		
couple	0.001	0.006	0.768	10.42	7.47	0.09		
ethnic origin[b]	0.270	0.003	0.067	1.22	8.70	3.35	18.1	3
student, employed	0.000	0.018	0.919	40.27	5.55	0.01	58.7	3
higher education	0.944	0.000	0.000	0.00	14.93	23.25	40.8	3
cultural participationc							24.4	9
both classical and popular	0.007	0.376	0.084	7.26	0.78	2.98		
classical only	0.046	0.015	0.002	3.98	5.86	9.88		
popular only	0.011	0.609	0.076	6.44	0.26	3.15		

a Reference category dependent: first contact at call 1 or 2, no refusal. reference categories 0. Chi-square 530 (df=39); Cox and Snell R^2 0.083, Nagelkerke R^2 0.092.

b Country of birth respondent, respondent's mother or respondent's father Morocco, Turkey, Surinam, Dutch Antilles or Aruba.

c Reference Family composition (adult(s) with children). Cultural participation (no participation).

Table 8.9 combines the different outcome types in which the respondents of the AVO can be subdivided. The dependent variable in this multinomial regression (see annex 7.1) comprises four categories: one or two calls to first contact and no refusal (easy respondents: the reference category, 54%), one or two calls and at least one (initial) refusal (23%), three or more calls to first contact and no refusal (17%), and finally the most difficult category: three or more calls to first contact and at least one (initial) refusal (7%). The first three columns present the β values for the three difficult groups compared to the easy reference group. As was to be expected, the results in table 8.9 are largely in line with those of the separate analyses of contactability and compliance in respectively table 7.5 and 8.7. With respect to urbanicity, the three difficult groups all differ from the easy respondents in roughly the same way. They less often have a listed telephone number (except for the converted refusals who were also hard to reach), more often live in a big city and less often in a detached dwelling and a well maintained neighbourhood and dwelling.

The three difficult groups do differ in their own specific way from the easy respondents. Among those who are hard to reach but not particularly reluctant to participate (the first column under β) singles and couples without children are overrepresented. They are underrepresented among those who cooperated reluctantly, and especially among the converted refusers who were easy to reach (second column under β). Ethnic minorities show a similar pattern as single person households. Persons who follow a study or are gainfully employed are overrepresented among the hard-to-reach, probably because their daily occupation takes them away from home, and slightly underrepresented among the easy-to-reach reluctant respondents. The better educated are underrepresented among the converted refusals, independent of the ease of contact. Cultural participation, regardless of the type, is higher among hard-to-reach respondents, and participation in classical cultural activities only higher among initial refusals. This effect occurs after inclusion of background variables related to cultural participation which indicates that, without response enhancing strategies, the outcomes on cultural participation would have been (slightly) biased.

These results confirm that there are different classes of participants. The model of classes of nonparticipants assumes that the reluctant respondents in table 8.9 are most similar to the final refusals. The largest effects in table 8.9 are related to urbanicity, family composition and sex. Urbanicity has extensively been discussed in section 8.2. Family composition and sex also were found to be a major determinant of reluctance in section 8.3.3. The next question is whether it can be assumed that the family composition and sex of the final refusals is comparable to the temporary refusals. This question will be addressed in the next section that focuses on the susceptibility of singles and males to refusal conversion.

8.3.3 *Does refusal conversion improve sample composition?*

Both section 8.3.1 and 8.3.2 have shown that among temporary refusals are relatively few single persons and men or, from a different point of few, refusal conversion brings in many extra women and families with children. This would improve data quality, if nonparticipation among these groups were high. If so, the conversion of temporary refusals would make the respondent group more similar to the net sample, and presumably the target population, and temporary refusals could be used a proxy for final refusals. This section will illustrate that refusal conversion in the AVO1999 did not have this desired, expected and beneficial effect.

Figure 8.4 Household composition respondents, initially and after subsequent inclusion of appointments, not able, refusal and population

Figure 8.4 illustrates the outcome of refusal conversion for a combination of sex and family composition. As the avo is a household survey, sex is only given for single-person households. A comparison with the Dutch population shows that, if only those respondents who immediately cooperated would be part of the survey, single males and three person and four person families would be underrepresented, and single females and two person families overrepresented. After following up on appointments and achieving interviews with those who were not able or not willing at the first contact, single females are included in proportion to their share in the population, families of three or more persons are slightly overrepresented, two-person families remain slightly overrepresented and single males are even more underrepresented than they were in the beginning. These results indicate that the negative effect for males and singles in the 'difficult respondent' categories is not so much because they are easy responders, but more probably because they are adamant refusers who simply do not allow themselves be converted.

Figure 8.5 **Single sample persons, age, sex and fieldwork effort (percentage of sex/age group among (1) easy respondent; (2) easy and hard-to-reach respondents, (3) all respondents (easy, hard to reach and initially reluctant) and (4) population)**

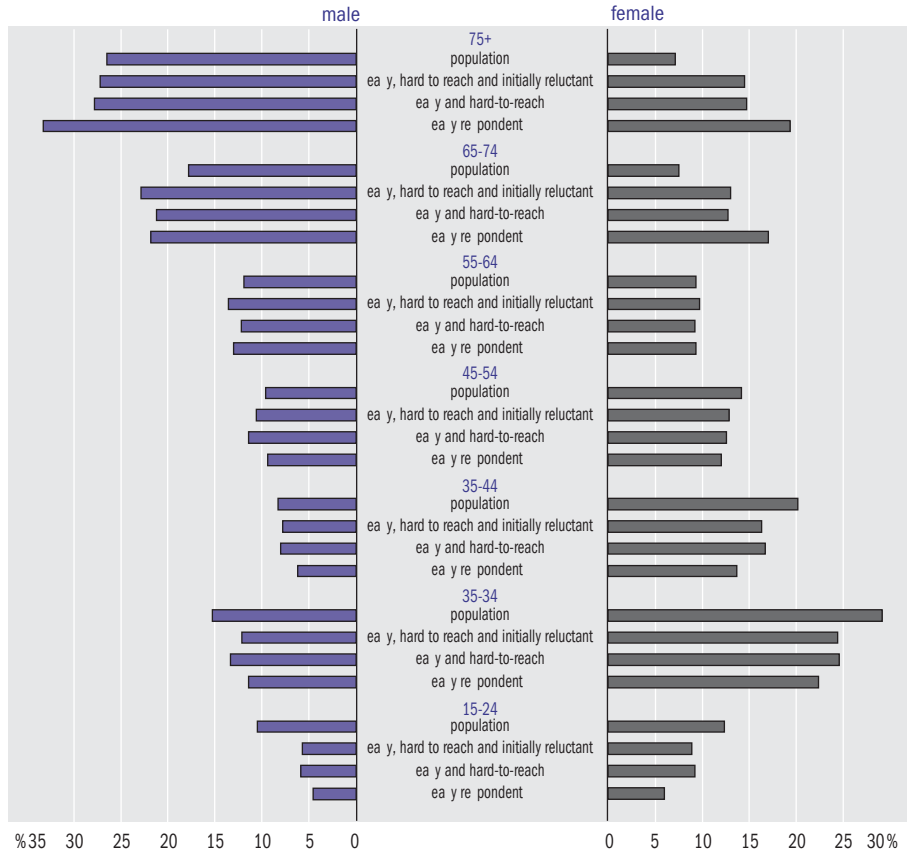


Figure 8.5 combines contactability and reluctance and shows the effect of the extended field efforts in the AVO1999 on the representation of single persons in the sample, subdivided according to sex and age. Younger single-person households are clearly underrepresented among the easy to reach, compliant respondents, and overrepresented among the older households. As can be seen, young singles are more often male, older singles more often female. Including hard-to-reach singles brings a slight improvement, as the overrepresentation of the elderly clearly diminishes. Converting refusals has a much smaller effect and more often negative than positive.

Singles and males may be more outspoken in their refusal. An alternative explanation is that they are more difficult to contact, and that there is simply less time available for additional contacts and refusal conversion after a late first, possibly temporary refusal. Later calls after a possibly temporary refusal may just result in

more noncontacts. To find out if this were true, the family composition of refusers should be known, which it is not. The interviewers did record the sex of the final refusers, and although there are approximately 25% missing values here, the available information shows that nonresponding men needed more calls to first contact than women (70% contact within two calls, compared to 77%) but also that after an unsuccessful first contact 61% of the nonresponding men and 62% of the nonresponding women received a visit from the interviewer when they were not at home. So it seems that the underrepresentation of males among difficult respondents is not due to that fact that women were easier to re-contact.

Figure 8.6 Percentage of respondents (6125), singles (1658) and single males (605) who immediately cooperated, by number of calls to first contact

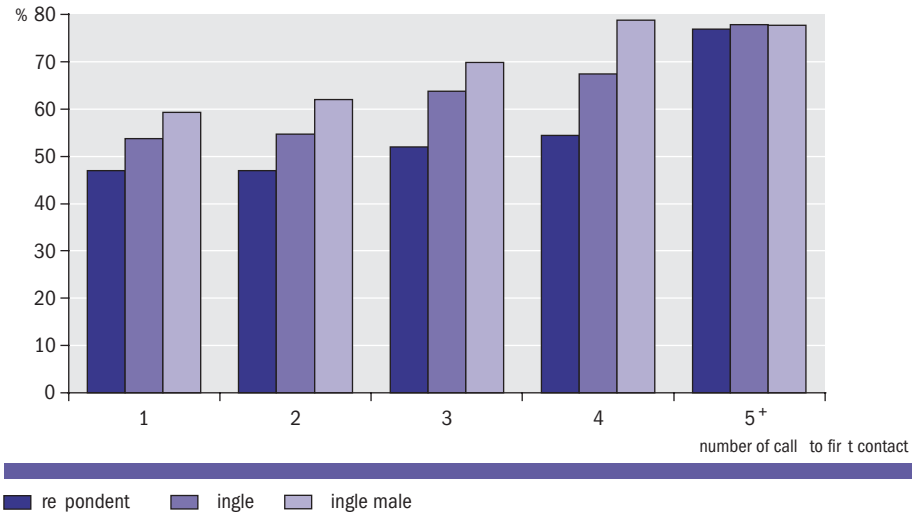
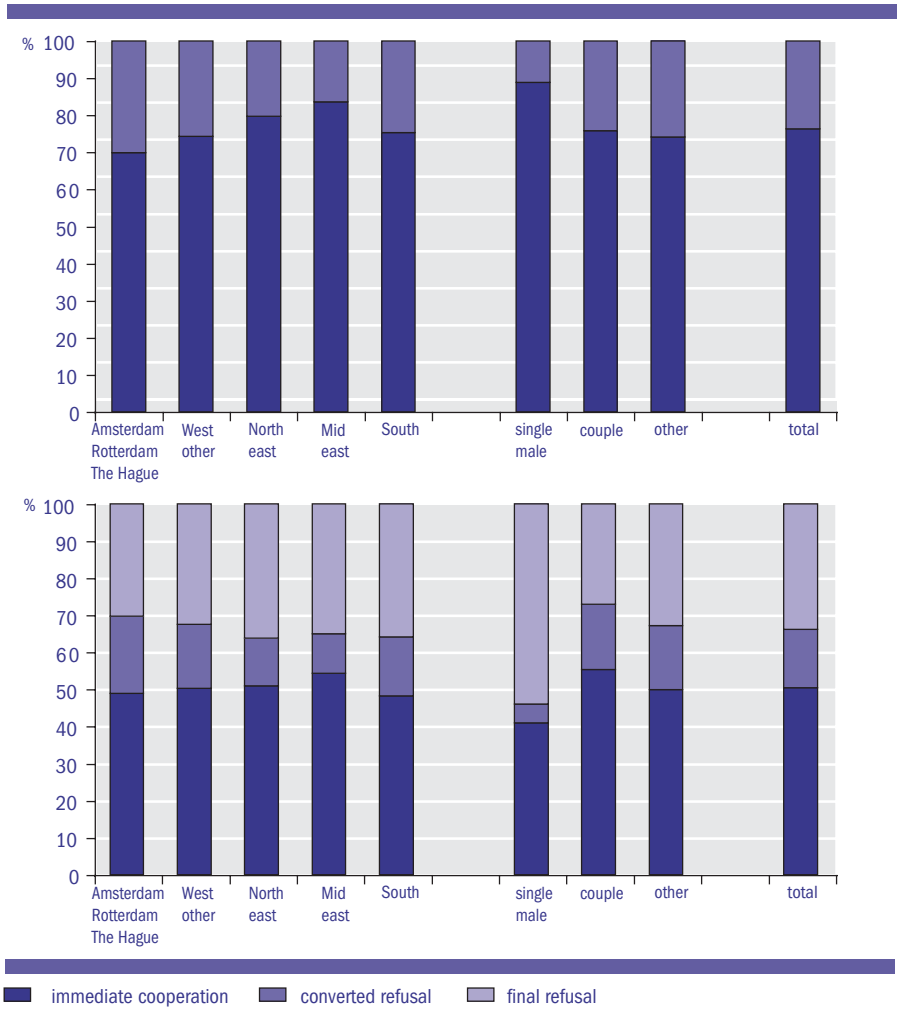


Figure 8.6 gives an additional illustration that this alternative explanation is not viable. It represents the relationship between contactability and reluctance for all respondents, single respondents and single males, and shows that, irrespective of the number of calls required to contact a household, responding singles more often cooperated immediately than all respondents, and single male respondents more than singles in general. This corroborates the idea that their underrepresentation among temporary refusals may be due to assertive refusal rather than cooperative behaviour, as especially male singles are noticeably underrepresented among the respondents. Only from call 5 on do differences between groups disappear. Figure 8.5 also shows that immediate cooperation is generally higher for late contacts, which confirms the suggestion that late contacts may frequently not get the chance to be converted and the only option for late contacts is immediate cooperation.

From the results presented in this section it appears that converting refusals in the AVO1999 yielded more respondents, including more males and singles, but was not particularly ade-

quate in getting the cooperation of those who were difficult to begin with. Put differently, converting refusals worsens the sample structure according to family composition and sex. A somewhat similar result can be observed when the education of the final respondent is taken into account. Statistics Netherlands data for 1999 show that 22.2% of the population of 15-64 years old had a higher education. Among the immediately cooperative respondents this was 23.4%, after interviewing those who needed an appointment for the interview 22.9% (closer), after converting those who were not able to cooperate 22.7% (even closer) and after including the converted nonrespondents who were initially not willing to cooperate 20.9% (alas). From this it appears that the interviewers did not succeed in converting highly educated sample persons. The difficulty of persuading the better educated to cooperate was also observed by the interviewers in the focus group discussion after the follow-up survey among persistent refusers (see section 6.3).

Figure 8.7 Con erted refusals as percentage of respondents (abo e) and percentage of net sample (below); region and household composition



Given this evidence it is not likely that temporary refusal are similar to final refusals, which is the idea behind the classes of nonparticipants model. Figure 8.7 illustrates the different ways in which sample persons can join the class of temporary refusals. The upper graph shows that among big-city dwellers are many temporary refusals and among single males few. The bar below shows that among big city dwellers there were many initial refusals and many households were converted. That leads to a sizeable number of converted refusals. Couples, on the other hand, were initially rather cooperative and if they refused, they were pretty susceptible to refusal conversion. Single males were initially very unwilling to cooperate and subsequently very unwilling to be converted. This interplay between initial cooperation and cooperation at a later stage works out differently for different groups. Figure 8.6 shows that it is unlikely that temporary refusals can represent final refusals. The main conclusion of this section is that indicators of reluctance are confounded with the survey process because it is very likely that 1) extra efforts are aimed at the least resistant temporary refusals, 2) field efforts are unequally distributed across easy and difficult geographical regions, and 3) different groups of (temporary) refusers differ in initial response rates and are not equally susceptible to refusal conversion techniques.

8.4 *Refusal conversion and data quality*

A recurrent concern regarding refusal conversion is that respondents who finally allow themselves to be persuaded end up delivering low-quality questionnaires. This concern seems less relevant for contacting hard-to-contacts. This group may yield data of different (poorer) quality, but if this is the case this will more likely be due to the different composition of late contacts than by them becoming irritated or showing satisficing behaviour (Krosnick, 1999) as they will be generally unaware of earlier calls.

Questionnaire quality can be measured by internal consistency, the presence and completeness of answers to open questions, the number of missing items, etc. In this section the relationship with reluctance and data quality is studied in terms of missing items. Firstly, the item nonresponse on the income question was studied. According to table 8.10 item nonresponse increases with the number of contacts up to 4, and then decreases. Those who immediately granted an interview and those who initially refused to cooperate have the lowest item nonresponse. Neither effect is significant. A logistic regression, yielding very low coefficients, shows that singles, the young and the old, and persons living in the poorer urban environments more often report their income. This suggests that income nonresponse is, if anything, more related to persons with higher incomes not willing to answer than to coercion of interviewers.

Income in table 8.10 represents family income and refers to a question in the face-to-face household interview. A second measure of item nonresponse is an overall

weighted count of item nonresponse for the complete paper-and-pencil questionnaire. Some item nonresponse is to be expected, especially as a number of questions were less relevant for elderly respondents. Inspection of the final paper questionnaires also showed, however, that a small number of respondents just picked the odd question and left entire pages blank. Figure 8.8 presents the results of a Multiple Classification Analysis of the item nonresponse count[4]. Each bar indicates the value on this indicator, after correcting for the other variables in the analysis.

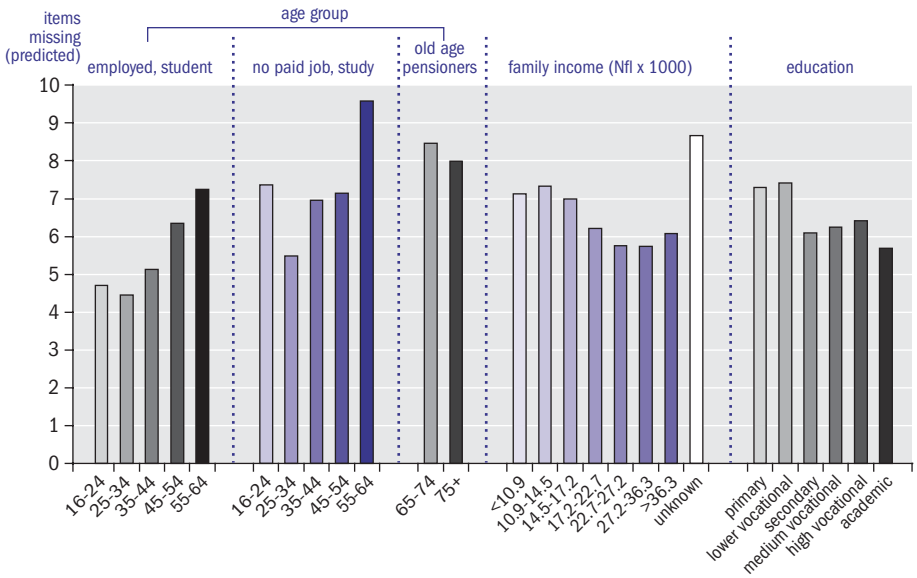
Table .10 Initial reluctance and income nonresponse

	% income missing	standard deviation	N
number of contacts			
1	8.5	27.9	3098
2	8.9	28.5	2067
3	10.5	30.7	724
4	11.2	31.6	188
5+	8.3	27.9	48
outcome first contact			
interview	8.5	27.9	3098
appointment, broken-off	9.8	29.7	1453
not able	10.8	31.0	362
refusal	8.6	28.0	1212
total	8.9	28.5	6125

The results indicate that item nonresponse increases with age (the exception being the youngest age group and especially those who do not follow a study and are not in paid employment), is lower for employed respondents and decreases with higher educational and income levels. In section 3.3.3 it was stated that forms skills or survey literacy may be related to response behaviour and that intellectual curiosity or cognitive skills might have a positive effect on survey participation. The present findings show that a similar effect may have occurred with respect to item nonresponse. Up and above the effect of these socio-demographic and socio-economic variables, item nonresponse differs for initially reluctant and cooperative respondents, although not in the expected direction. Respondents who were not able or willing to be interviewed at the first contact but could be persuaded to cooperate after all, actually do have a lower nonresponse rate than those respondents who immediately cooperated by granting an interview or making an appointment. The effects are rather small but give no indication at all that refusal conversion will lead to poor data quality.

4 Multiple Classification Analysis is an extension of analysis of variance. It computes the predicted mean of the outcome variable (here income nonresponse) and the deviation to the grand mean for each level of each factor in the ANOVA adjusted for the other factors.

Figure 8.8 Predicted eighted count of item missings, AVO 1999



8.5 *Conclusion*

This chapter on reluctance to cooperate has aimed at identifying those who are less likely to cooperate, showing how cooperation rates can be substantially increased by converting initial refusals, measuring the degree of willingness or the type of reluctance of sample households and persons, finding out if enhancing response rates increases accuracy and assessing whether winning over reluctant respondents results in incompletely filled in questionnaires.

To identify those groups that are less likely to cooperate, information is required on reluctant respondents and on final nonrespondents. Data on the latter group are available from the follow-up survey among persistent refusals, to be presented in the next chapter. In the present chapter only very limited information was available on final nonrespondents: from sample and population data the family composition and urbanicity of final refusals is known or can at least be approximated. Complete information is, of course, only available for the most willing part of the sample (the respondents), which makes it possible to relate the willingness to cooperate of respondents to background and survey variables. From the population and sample data and the contact forms it can be deduced that big city dwellers were initially reluctant but could be persuaded to cooperate at later contacts. Without refusal conversion attempts, big-city dwellers would have been characterized as less likely to cooperate; after refusal conversion they were as least as likely as inhabitants from more rural areas to cooperate (see figure 8.2). The results for family composi-

tion show an entirely different pattern. There were relatively few singles and males among the converted refusals. Rather than indicating that these groups were more likely to cooperate, the conclusion should be that they less frequently cooperated immediately, and even less frequently after refusal conversion, and were thus doubly uncooperative. These results indicate that basing conclusions about survey cooperation on (cooperative and reluctant) respondents merely leads to erroneous outcomes if reluctant respondents are not similar to final respondents, which was the case for family composition. Given this limitation, one interesting result from the analysis of reluctance is that participation in classical cultural activities is only higher among converted refusals, which suggests that they do not conform to the stereotype of socially isolated, or socially non-participative persons who are also unlikely to take part in a survey.

The question of who is likely to refuse can thus be only very partially answered with the information on respondents analysed in this chapter. On the other hand, clear evidence is presented on how a thorough refusal conversion strategy can substantially enhance response rates. If only those had been interviewed who complied at the first contact (interview, appointment, broken off and taken up again) the response rate would have been about 50% instead of the final 66%. About half of those who were ill, entertaining, too busy or otherwise engaged cooperated at a subsequent contact, as did about a third of the initial refusals.

Measuring the extent of the reluctance to cooperate is much more complex than measuring contactability (section 8.1). In this chapter three indicators of reluctance were discussed, namely the reason for noncooperation, interviewer' assessments of the likelihood of future success after a refusal and finally the number of (temporary) refusals. Section 8.2.1 showed that reasons for noncooperation may only partially reflect why persons do not comply, because unwilling sample households gave different reasons at subsequent contacts. For instance, of those who initially said they were too busy to participate more than a third plainly refused at a subsequent contact. Saying "too busy" or "no time" may just be an easy way out to avoid lengthy discussions with the interviewer on reasons for refusal or arguments to cooperate. Interviewer' assessments of the likelihood of future cooperation of initial refusals are also an imperfect indicator of reluctance (see section 5.2). Table 8.4 showed that those who were characterized as absolutely non-cooperative were re-approached in 70.5% of the cases, of whom almost half finally cooperated. Those who appeared to be less forbidding were re-approached in 77.5% of the cases, of whom somewhat more than half cooperated, which are small differences. In addition, the results from table 8.8 suggest that interviewers in big cities may have been less likely to classify a refusal as absolutely impossible of being a future success. The final measure of reluctance, the number of (temporary) refusals or unsuccessful calls, is also inadequate as the selection of which interviews to re-issue and the willingness to cooperate at later contacts are confounded. Table 8.3 illustrated how interviewers

tend to go back (and back again) to promising cases only. These findings suggest that there is no readily available single measure of reluctance of both respondents and nonrespondents, and that different indicators of reluctance should preferably be compared or combined.

In section 5.2 two models were introduced, the *continuum of resistance model* and the *classes of nonparticipants model*. The continuum of resistance model, when applied to willingness to cooperate, assumes that final nonrespondents are similar to the most reluctant respondents and that there is a linear correlation between reluctance and survey variables. The present data give ample reason to reject this model for reluctance, even if no survey variables are available here for final nonrespondents. The absence of a clear, one-dimensional measurement model of reluctance, and the fact that the family composition from converted refusals mirrors the family composition of final refusals, shows that there in no continuum of reluctance (table 8.8).

The model in section 8.3.1 can also be considered as a classes of participants model if it is not assumed that the response groups indicate an increasing level of reluctance but different types of reluctance. In this case the different types of participants (immediately cooperative, appointment first, resumed interview, cooperation after initial 'not able', converted refusals) would each have to be similar to types of final nonparticipants (broken-off, not able, refusal). This model is also difficult to test here, because very little information is available on final nonparticipants (except from the follow-up study to be discussed in the next chapter) and because, according to table 8.1, the vast majority of the nonrespondents belong to two groups only: incomplete sets and final refusals. What can be concluded, if the groups identified in table 8.8 are interpreted as classes of respondents, is that temporary refusals differ from those who were initially not able, who needed an appointment or who participated immediately.

The classic classes of nonparticipants model would assume that hard to reach respondents are similar to noncontacts and initially reluctant respondents to final refusals. In practice, respondents that are easy or difficult to contact and willing or reluctant to cooperate are often compared, as was done in section 8.3.2 in which accessibility and amenability have been combined. The results in table 8.9 confirm the earlier results of previous analysis of contactability and cooperation from chapter 7 and chapter 8. This analysis did not include nonparticipants, however, so it was not a proper test of the classes of nonparticipants model.
Section 8.3.3 investigated whether temporary refusals are similar to final refusals. This is a very important question, because if temporary refusals are (more) similar to final refusals, the accuracy of survey estimates will improve as a result of refusal conversion and the survey outcomes of temporary refusals can be used as proxy for the final refusals. However, as mentioned above, all evidence in this section indicates that in this study temporary refusals are not at all similar to final refusals. This

means firstly that neither the continuum of resistance model applies, nor that the class of reluctant participants is similar to the class of refusing nonparticipants. It also means that refusal conversion in this case is not likely to have led to more accurate survey estimates as the final sample of the AVO1999 is less similar the population after refusal conversion than before. And finally it means that it is very doubtful whether refusal conversion reduced nonresponse bias. This is the subject of the next chapter, where survey information on final nonparticipants will be incorporated.

One question remains, namely whether refusal conversion has a detrimental effect on survey quality. This was explored in section 8.4, and from the available evidence there is no indication that this is the case.

9 Are refusers different?

9.1 Introduction

In chapter 6 the design and the fieldwork of the AVO survey were presented, chapter 7 investigated how hard-to-contact respondents differed from easy-to-contact respondents, and chapter 8 looked at how reluctant respondents differed from compliant respondents. This chapter explores how final refusals differ from respondents and whether extended field efforts manage to get difficult respondents on board that are similar to final refusals. For this purpose a subsample of persistent refusals in the AVO1999 survey were re-approached in a follow-up survey. Chapter 6 showed how high quality (and high cost) fieldwork procedures and strategies managed to secure the cooperation of the majority of these persistent refusals. Section 9.2 provides information on how well the gross sample of the follow-up survey appears to represent the total nonresponse group from which it was drawn. This section also discusses the dropout from the follow-up sample due to ineligibility and non-contact and the differences between participants and refusers in the follow-up survey, indicates how the inclusion of multiple modes in the follow-up survey may have contributed to its success, and presents the control group that was introduced to compensate for possible consequences of the adapted design of the follow-up survey.

Section 9.3 first presents univariate outcomes comparing the AVO and the follow-up survey and, when available, all nonrespondents and the control group that was introduced to control for design effects (section 6.3.1). Three types of variables will be presented: variables related to (social) environment, background characteristics of the respondent and the household and substantive variables that were key variables in the AVO, that may reflect social involvement (see section 4.2.1) and that showed significant effects in chapters 7 and 8 or separated respondents from refusers. Subsequently a multivariate model is presented in section 9.3.2 that explores how final refusers differ from different classes of participants. This can be seen as a test of the classes of nonparticipants model presented in chapter 5. Section 9.4 tests the stability of these outcomes, which are based on the small sample of persistent refusers only, by presenting the results of a sensitivity analysis and the results of a randomization study. In section 9.5 the answers to the main questions of this study are addressed, based on the empirical evidence presented in chapters 7, 8 and 9, namely whether a follow-up study among persistent refusals is a viable way of collecting information with which to measure nonresponse bias, whether temporary refusals are similar to final refusals, and whether higher response rates lead to more accurate survey estimates.

9.2 Results follow-up survey

9.2.1 Sample composition

The follow-up survey, introduced in section 6.3.1, was based on a sample of 350 households drawn from the 1756 more persistent of the 2521 refusals. This may have minimized the risk that a larger part of the sample would consist of sample units that did not participate in the regular survey for superficial reasons such as bad timing, or a high initial response rate in a particular region and consequently less ardent field efforts. The drawback is that this procedure may have delivered a gross sample for the follow-up survey that is dissimilar from the total group of nonrespondents. This drawback may be exacerbated by the small size of the follow-up sample. Therefore, before comparing the follow-up survey and the regular survey, an assessment should be made of how similar the sample drawn from the persistent refusers is to all nonrespondents, apart from being slightly more 'difficult' than in general. In addition, an investigation is needed into whether there is a difference between persistent refusers who cooperated in the follow-up survey and those who did not. These questions are difficult to answer as there is not much information available either on the nonrespondents in the regular AVO or on the nonrespondents on the follow-up survey. This section will try to address these questions, explain the focus on cooperation rather than on response, and describe the control group.

Gross sample and other nonrespondents

An analysis of neighbourhood characteristics showed that there is hardly any difference between the follow-up sample households and the other 2786 nonrespondents. None of the indicators of urbanicity differed significantly between the selected and the non-selected. Only one variable distinguishes between these groups: a listed phone number was available for 82% of the households selected for the follow-up survey and for only 74% of the other nonrespondents. From the 70% of cases where an estimated age of the refuser – as recorded in the regular AVO – is available, it turns out that the follow-up sample is approximately 50.1 years old, and the non-selected nonrespondents 49.8 years (see table 9.1).

Table 9.1 ge and response groups[a]

	questionnaire	estimate interviewer	% estimate available	N[b]
respondents	49.3	49.1	77.1	6125
nonrespondents		49.8	68.2	3136
not selected follow-up survey		49.8	67.6	2786
sample follow-up survey		50.1	73.1	350
contacted		50.4	73.9	329
cooperation	49.6	48.8	72.8	235
refusal		54.2	76.6	94
total	49.4	49.3	74.1	9261

a Correlation estimated age and age from questionnaire .91 (N=4887)
b 6125 respondents and 3136 nonrespondents add up to the 9261 households accounted for in table 6.3.

An inspection of fieldwork records indicates that those selected for the follow-up survey are indeed more 'difficult' than the other nonrespondents. Table 9.2 compares fieldwork information from the regular survey for those nonrespondents who were re-approached in the follow-up survey, the non-selected nonrespondents and the total group of nonrespondents. The table shows that those who were selected for the follow-up survey had refused 1.57 times on average in the regular AVO, compared to 1.33 times for all nonrespondents. Their non-cooperation was less often due to not being able (0.11 times, compared to 0.2 times for all nonrespondents). The table also shows that fewer than one per cent of the selected nonrespondents never refused in so many words, whereas for more than 10% of all nonrespondents there never was an actual refusal recorded. In 73% of the cases the interviewer in the regular AVO assessed that this household absolutely did not want to be interviewed, compared to 45% of all nonrespondents.

Table 9.2 Indicators of reluctance among nonrespondents in regular AVO for those selected for followAup survey and other nonrespondents

	number of calls in which reluctance was shown		reluctance					assessment future success	
	not willing	not able	0	1	2	3	4+	absolutely not	N
selected in follow-up sample	mean	mean	%	%	%	%	%	%	
selected	1.57	0.11	0.6	59.7	23.7	14.9	1.1	73	350
not selected	1.30	0.21	12.1	55.8	22.6	9.1	0.5	41	2786
total	1.33	0.20	10.8	56.2	22.7	9.7	0.5	45	3136

The following reservations should be made with regards to these results. Firstly, although the follow-up sample seems to be a particularly tough group, the actual firmness of the refusal always rests on subjective assessments, as was indicated by table 8.4 where it is shown that even if an interviewer considered an interview to be absolutely impossible in the future, many refusers could be converted. Secondly, recorded reasons for refusal were used in the selection of the follow-up sample, but were not keyed (see section 6.3.3). Thirdly, it is unknown who within the household was the refuser at subsequent contacts (although every household member ultimately had to cooperate). And finally, persistence in refusing is a double-barrelled indicator, as shown in chapter 8. Adamant refusers in many cases refuse only once, whereas a high number of refusals may indicate that the fieldwork organization and the interviewer persist in feeling that at the end of the day the reluctant household will cooperate. Nonetheless, it seems that those who were sampled for the follow-up survey are by no means the fairly easy nonrespondents, but are probably more correctly designated as tough nonrespondents.

Ineligibles and noncontacts

A number of sample households in the follow-up survey will be excluded from further study. One address turned out to be faulty, two households had moved and four addresses appeared to be uninhabited. In another case, lack of knowledge of the native language turned out to be a problem. At 13 addresses nobody was found at home, despite frequent calls and despite the fact that half of these the residents were contacted at the first two calls in the regular AVO. It is possible that these households had moved and that the dwelling was empty, that they were away on long vacations, or could not be contacted for other reasons. As in the regular AVO, no independent information on empty dwellings was collected (see section 6.2.2). In the end, 21 addresses were classified as ineligible or could nor be contacted. As the noncontact rate in the AVO did not seem to be a particular problem (smaller than 2%), and the follow-up survey was especially focused on refusals, in this chapter only the 329 contacted households in the nonresponse survey will be taken into account. This means that the main emphasis is on cooperation in the follow-up survey, rather than on response.

Refusers and participants in the follow-up survey

If there were large differences between the refusals and the respondents on the follow-up survey, this would limit its usability. An analysis of differences between respondents (235) and refusals (94) in the follow-up survey did not show any relationship with urbanicity at all, except for the proportion of ethnic minorities in the neighbourhood.[1]

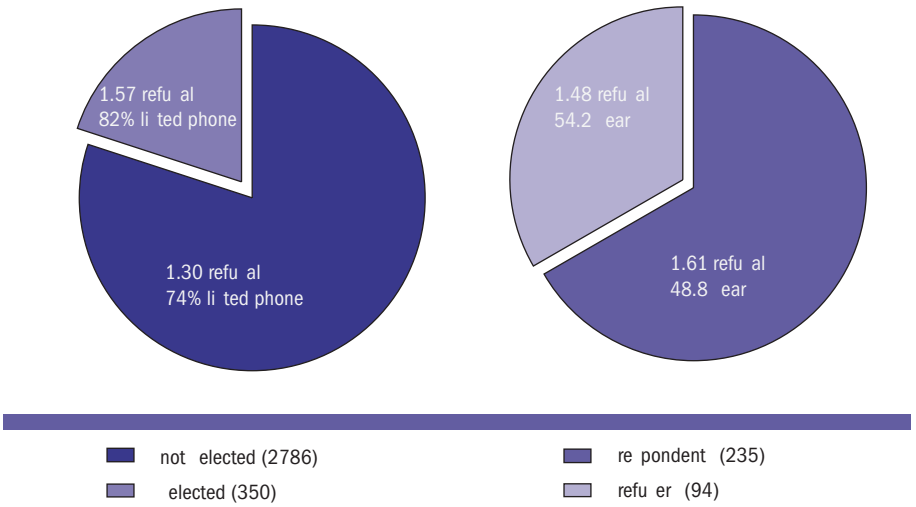
1 The original classification for ethnic minority rate was (1) 0% (2) 1-5% (3) 6-10% (4) 11-15% (5) 16-25% and (6) 26+%. A numerical transformation was based on class means: (1) 0% (2) 3% (3) 8% (4) 13% (5) 20.5% and (6) 30%. The original category scores yielded a significant difference between regular and follow-up survey, the numerical estimates did not.

In the entire AVO sample this was 8.94%, among the hard-to-contact respondents (10.44%) and among the cooperative, easy to contact respondents (7.80%). The average for the total follow-up sample (N=350) was 8.96%, for the participants in the follow-up survey (8.96%) and for the refusals (7.70%). An inspection of the family names of the participants in the follow-up survey revealed mainly Dutch names, and very few of foreign descent. Schouten, Bethlehem and Stoop (2002a) suggested a probable underrepresentation of ethnic minorities in the AVO after a comparison with population totals. This cannot be explained by the relatively small number of non-Dutch speakers who were excluded from the sample (see table 6.3), however, nor is there any evidence that ethnic minorities are underrepresented among the participants in the follow-up survey.

Both respondents and refusals in the follow-up survey had refused at least once in the regular AVO, namely 1.61 times (follow-up participants) and 1.48 times (follow-up refusals). This difference is mainly due to the fact that 57% of the follow-up respondents refused only once in the AVO, compared to 66% of the refusers in the follow-up survey. These results may be seen as another illustration of the complicated relationship between number of contacts (or refusals) and reluctance: the most difficult refusers may be those who refuse only once, and are so outspoken that there seems to be no use in re-contacting them.

A final difference between respondents and refusers in the follow-up survey is age. According to the estimated age in table 9.1, the respondents (48.8 years) are substantially younger than the refusers (54.2 years). According to this table 9.1 the estimated age seems to be a reliable source of information, as it is remarkably similar to the recorded age for respondents. It is not known, however, whether interviewers may have adapted their answer to the information they obtained from the questionnaire. The age difference between respondents and nonrespondents on the follow up survey will come back in section 8.3.2. A graphical representation of the selection of the follow-up sample and the final cooperation is given in figure 9.1. In the remainder of this chapter the 235 respondents on the follow-up survey will generally be considered to represent all refusers in the regular AVO.

Figu e 9.1 Selection follow-up sample from refusals AVO (left) and participants and refusals follow-up survey (right)



Control group

The design of the follow-up survey differed from that of the regular AVO. The questionnaire was much shorter, the mode of interviewing differed, only one person had to participate, questions had been adapted and the period of interviewing was June-August 2000 as opposed to September 1999-January 2000. To control for these effects a fresh sample was drawn of 300 addresses. After the removal of ineligibles, 276 households remained, from which a response rate of 64% was achieved (N=177). The response rate would have been much higher but for the relatively high noncontact rate (15%, compared to less than 2% in the regular AVO). As this group was only to be used as a control for design effects, a high or low response bias did not seem particularly important. The results from the control group will be discussed in the next section, where a univariate comparison will be presented between respondents on the regular AVO and the refusers who responded in the follow-up survey.

9.2.2 *Interview mode*

In order to enhance response rates, respondents in the follow-up survey could choose from different interview modes. Table 9.3 presents the outcome of this selection. The table also contains an extra category: being interviewed by the central telephone unit of the organization. This unit telephoned nonrespondents, initially with the purpose to ask for reasons for refusal. In the end 16% of the successful interviews were conducted by this unit.

The number of respondents in the follow-up survey is very small. Relationships between respondent characteristics and preferred mode are generally not significant. Nonetheless, the results in table 9.3 and the outcomes of the multinomial regression in table 9.4, with interview mode as dependent variable, suggest that a choice of interview mode may increase response rates, especially in big cities and of busy people, namely those who are in (paid) employment, are a student or participate in cultural activities. More than half of the 25 big-city dwellers who finally cooperated, were interviewed by telephone, compared to less than one quarter of all respondents. Face-to-face surveys seem to be ill-suited to those who participate in popular culture (going to the cinema, pop concerts, dance events): less than half of those who participated in both popular and classical cultural activities and 60% of the popular culture participants were interviewed in the standard face-to-face way. Other variables, such as family composition, made little difference.

Two conclusions can be drawn from the results presented in tables 9.3 and 9.4, with all due reservations concerning size, the peculiar composition of the follow-up sample, and uncertainty about what exactly went on in the field. Firstly, it seems that offering alternative modes is most effective for those groups that are similar to late contacts rather than to reluctant respondents. Secondly, it turns out again that big-city dwellers, this particularly fearsome group of nonrespondents, will – with some extra efforts – cooperate to the same extent – or even more – than the inhabitants of regions that are generally seen as more accessible and more amenable.

Table 9.3 Interview modes in followAup survey

	interviewer				telephone unit			
	face to face	tele-phone	self-com-pletion[a]	refusal	suc-cess	refusal	total	N
Amsterdam, Rotterdam, The Hague	26	20	9	6	17	23	100	35
randstad, other	44	2	9	7	11	26	100	88
north east	60	2	4	14	0	20	100	50
mid east	58	0	3	4	12	22	100	67
south	47	7	9	6	15	17	100	89
total	48	5	7	7	11	22	100	329
Amsterdam, Rotterdam, The Hague	36	28	12		24		100	25
randstad, other	66	3	14		17		100	59
north east	91	3	6		0		100	33
mid east	80	0	4		16		100	49
south	61	9	12		19		100	69
cultural participation								
both	47	16	13		24		100	62
classical only	77	6	6		10		100	31
popular only	60	2	17		21		100	52
none	83	3	4		9		100	90
employed, student	55	8	14		22		100	119
other	80	5	5		9		100	116
total	68	7	10		16		100	235

a Questionnaire collected by interviewer (21) and Internet version (2)

9.3 *A comparison of refusals, respondents and other groups*

9.3.1 *Socio-structural characteristics and survey outcomes*

Tables 9.5, 9.6 and 9.7 compare the results of the regular AVO respondents, the refusers who participated in the follow-up, and the control group. The participants in the follow-up survey will be called '*the refusers*' here and should not be confused with the 'other' nonrespondents in the AVO (nonparticipants who did not take part in the follow-up survey) nor with those who refused to participate in the follow-up survey (double refusers). Table 9.5 includes information on the AVO-respondents, the refusers who participated in the follow-up survey and the other nonrespondents, and compares groups according to region, neighbourhood and dwelling, available from regional databases and interviewer observation. No interviewer observation is available for the control group. Table 9.6 presents socio-structural characteristics of the participating households, and table 9.7 substantive survey outcomes. Two selection criteria were used for inclusion in the latter table. Firstly, variables had to measure social or cultural participation, as this was both the subject of the surveys and is generally assumed to be related to response behaviour (see section 4.2.1).

Table 9.4 Multinomial regression interview mode followAup survey[a]

	β	exp β	standard error	Wald	sign.
telephone					
intercept	–4.139		0.738	31.46	0.000
Rotterdam, Amsterdam, The Hague	2.923	18.591	0.694	17.75	0.000
cultural participation					
both	2.403	11.062	0.785	9.38	0.002
classical	0.682	1.977	1.011	0.45	0.500
popular	–0.218	0.804	1.231	0.03	0.860
employed, student	0.433	1.542	0.635	0.46	0.495
self-completion					
intercept	–3.483		0.600	33.65	0.000
Rotterdam, Amsterdam, The Hague	1.091	2.978	0.748	2.13	0.144
cultural participation					
both	1.474	4.365	0.670	4.84	0.028
classical	0.431	1.538	0.905	0.23	0.634
popular	1.378	3.965	0.665	4.29	0.038
employed, student	1.020	2.772	0.529	3.72	0.054
telephone unit					
intercept	–2.777		0.445	39.00	0.000
Rotterdam, Amsterdam, The Hague	1.356	3.881	0.603	5.05	0.025
cultural participation					
both	1.460	4.307	0.512	8.14	0.004
classical	0.149	1.161	0.727	0.04	0.837
popular	0.936	2.549	0.540	3.00	0.083
employed, student	0.897	2.452	0.421	4.54	0.033
model	χ^2		df		
Rotterdam, Amsterdam, The Hague	18.67		3		0.000
cultural participation	25.52		9		0.002
employed, student	7.43		3		0.059
total	59.46		15		0.000

a Dependent: interview mode (reference group: face-to face).
Reference categories (=0) Other regions, no cultural participation, not employed, student.
Cox and Snell Pseudo R^2 = 0.224, Nagelkerke Pseudo R^2 = 0.262.

Table 9.5 Region, neighbourhood and dwelling avo response, refusers (in followAup survey), other nonrespondents and control group (eligible addresses) %[a]

	VOA response	followAup refusers	other nonA respondents	total sample	control
region					
Amsterdam, Rotterdam, The Hague	16.5	10.6	14.2	15.6	11.9
randstad other	29.5	25.1	27.8	28.8	20.9
north east	10.4	14.0	11.3	10.8	15.8
mid east	20.3	20.9	21.4	20.6	24.3
south	23.4	29.4	25.4	24.2	27.1
N	6107	235	2892	9234	177
high-rise buildings					
yes	33.7	32.8	38.1	35.1	31.6
no	66.3	67.2	61.9	64.9	68.4
N	6125	235	2901	9261	177
ethnic minorities					
none	12.3	8.9	11.9	12.1	14.1
1-5%	34.9	36.2	35.0	34.9	33.9
6-10%	21.6	25.5	21.1	21.7	27.7
11-15%	13.4	10.6	13.4	13.3	13.6
16-25%	10.5	12.3	10.9	10.6	5.6
26+ %	7.4	6.4	8.0	7.5	5.1
N	6122	235	2901	9258	177
age neighbourhood					
<1946	22.3	24.7	23.6	22.8	
1946-1974	37.9	37.7	37.7	37.8	
>1974	39.8	37.7	38.7	39.4	
N	5873	223	2745	8841	
maintenance neighbourhood					
excellent	20.8	21.3	20.1	20.6	
good	63.4	61.8	60.5	62.5	
mediocre	14.6	14.7	17.8	15.6	
poor	1.2	2.2	1.6	1.4	
N	5930	225	2772	8927	
location neighbourhood					
in city centre	8.9	9.0	9.4	9.0	
close to centre	33.6	31.8	31.7	33.0	
outside centre	57.5	59.2	58.9	58.0	
N	5841	223	2754	8818	
type of dwelling					
detached	14.1	15.1	16.3	14.8	
semi-detached	25.7	20.0	18.9	23.5	
terraced house	31.6	34.2	30.1	31.2	
multi-unit	26.5	27.6	30.5	27.7	
other	2.1	3.1	4.1	2.7	
N	5963	225	2766	8954	

Table 9.5 (cont.) **Region, neighbourhood and dwelling AVO response, refusers (in follow-up survey), other nonrespondents and control group (eligible addresses) %[a]**

	VOA response	follow-up refusers	other non-respondents	total sample	control
maintenance dwelling					
excellent	23.9	22.2	21.7	23.2	
good	61.2	62.4	60.0	60.9	
mediocre	13.3	14.5	16.0	14.1	
poor	1.6	0.9	2.3	1.8	
N	5867	221	2720	8808	

a Significance differences at 0.01 - level

AVO-response – follow-up refusers:	Region
AVO-response – other nonrespondents:	All variables except Ethnic minorities (n.s)
AVO-response – control group:	Region
Follow-up refusers – other nonrespondents:	None
Follow-up refusers – control group:	None

And secondly, only those variables were included in the table, and in future analyses, that showed significant differences between groups of respondents and between respondents and refusers. This implies that the major part of the variables did not show significant differences.

A complicating factor in comparing groups is the differential rate of missing items across variables. In general, the item nonresponse was lower in the follow-up survey and in the control group. This can be chalked up to the substantially reduced length of the questionnaire and to the predominant interview mode in the follow-up survey. Whereas most questions were self-completion in the regular AVO and most respondents did skip some items, the interviewer in the follow-up survey were apparently much more conscientious in filling in all questions. There is, however, one exception. In the regular AVO the importance of obtaining income data was urged upon the interviewers, as this was one of the key variables of the survey. In the follow-up survey and the control group, respondents were told that they could skip questions they did not want to answer, which brought about a high rate of missing incomes. In table 9.5, 9.6 and 9.7 percentages are computed across non-missing values, and for income the percentage of item missing in given in addition.

Table 9.5 presents the data on region, neighbourhood and dwelling for respondents and nonrespondents in the AVO, the refusers and the control group. It is immediately clear that there are no systematic differences between the groups. The refusers are less often big city-dwellers which was to be expected as field efforts in the regular survey already managed to convert relatively many city dwellers. Besides, the group of big city-dwellers in the follow-up sample is so small (35) that differences are not significant. A comparison between dwelling type shows that AVO respondents more

often live in semi-detached houses, and refusers more often in terraced houses, whereas the percentage in multi-unit buildings and in detached houses is remarkably similar. All in all, the differences are amazingly small and neither of them is significant.

Table 9.6 shows that difference in socio-structural characteristics between the AVO-respondents, refusers and control group are generally insubstantial. Compared to the AVO-respondents, among the refusers are more males. In the regular AVO, however, the entire family had to cooperate after an introductory face-to-face interview with a knowledgeable adult, so the sex of the respondent would only be an issue for single males. The latter, and single females too, are more present in the follow-up survey than in the regular survey, which was to be expected as chapter 8 showed that they were seriously underrepresented in the AVO. There are no differences in average age between AVO-respondents and refusers (see table 9.1). The latter less often belong to the middle group, and are more often either somewhat younger or somewhat older than the AVO-respondents. Differences with regard to ethnic background are negligible, and with regard to social group small. Among the refusers seem to be fewer lower educated persons (primary education only) and fewer higher educated persons (academics). Home ownership makes no significant difference. The comparison of household income is difficult because, as explained above, the follow-up group contains substantially more missing data. For those households whose income is available, it seems that the follow-up group is less affluent than participants in the regular AVO (55% in the lower three income brackets, compared to 44% of the AVO-respondents). Surprisingly, one of the largest differences occurs in a soft demographic variable: religion. 43% of the refusers say they do not have a religious affiliation, compared to only 35% of the regular respondents.

Table 9.6 ackground characteristics respondents, followAup group, control group %[a]

	VO response	followA up	conA trol		VO response	followA up	conA trol
sex				education			
male	37.1	41.3	41.8	primary	20.2	15.4	20.0
female	62.9	58.7	58.2	lower vocational	20.2	23.2	17.7
N	6125	235	177	secondary	19.3	21.5	15.4
				medium vocational	21.4	22.8	22.9
family composition				high vocational	13.5	14.0	17.1
single male	9.9	12.8	12.4	academic	5.5	3.1	6.9
single female	17.2	20.0	15.3	N	5943	228	175
couple	34.6	35.7	31.1				
couple with children	33.4	28.5	34.5	home ownership			
single parent	4.9	3.0	6.8	owner	54.8	51.5	56.5
N	6125	235	177	tenant	45.2	48.5	43.5
				N	6125	233	177
age							
16-24	3.7	4.8	2.3	religion			
25-34	19.5	20.7	18.2	none	35.4	43.4	33.3
35-44	21.0	22.5	23.9	reformed, calvinist	24.5	20.4	27.1
45-54	19.2	14.1	18.2	roman catholic	33.0	29.8	31.1
55-64	13.9	12.8	14.8	other	7.1	6.0	7.9
65-74	13.4	15.0	11.9	N	6041	235	177
75+	9.2	10.1	10.8				
N	6125	227	176	social group			
				employed	43.9	40.9	47.7
ethnic group				self-employed	4.7	7.7	5.7
Turkey, Morocco Surinam, Antilles	2.5	2.1	3.4	unemployed	10.9	6.8	9.1
Dutch, other	97.5	97.9	96.6	disability pension	5.5	7.2	3.4
N	6125	235	177	old age pension	20.6	22.6	21.6
				student	1.5	2.1	0.6
number of earners				homemaker	12.8	12.8	11.9
single person, no job	16.3	19.1	16.9	N	5991	235	176
single person, job	10.7	13.6	10.7				
no earner	18.4	19.1	16.4	income (missing excl.)			
single earner	21.8	19.1	22.0	< 10.9	14.0	15.3	16.4
multiple earners	32.7	28.9	33.9	10.9-14.5	15.6	21.2	17.1
N	6125	235	177	14.5-17.2	14.2	18.8	12.9
				17.2-22.7	17.5	15.9	13.6
income				22.7-27.2	13.5	10.6	13.6
known	91.1	72.3	79.1	27.2-36.3	15.3	10.0	14.3
unknown	8.9	27.7	20.9	> 36.3	9.9	8.2	12.1
N	6125	235	177	N	5577	170	140

a Significant differences
AVO-response – follow-up refusers: Age, homeownership, income missing (0.01)
Family income, number of earners (0.10)
AVO-response – control group: Age, income missing (0.01)

Table 9.7 gives an overview of survey variables that distinguish between the AVO-respondents, the refusers and the control group. Here too, differences are generally small, or not significant. Health, surprisingly, distinguishes the control group from the other ones, basically because they are significantly less often very healthy. Refusers less often have a family member or close friend who is chronically ill and needs help, and – if they do – less often take care of these persons. Due to the small sample size, however, it will be difficult to generalize this result. The pattern of cultural participation of the refusers differs from the regular AVO and the control group. They more often attend no cultural performances at all, less often attend classical culture performances (theatre, classical concerts, opera), are less often indiscriminate art lovers (both classical and popular) but more often go to pop concerts, dance events and the cinema. For instance, 13.6% of the AVO respondents go to the cinema at least once a month and 17.4% of the refusers. The higher rate of item missing in the AVO (2.2% compared with 0% in the follow-up survey) is not sufficient to explain this difference. The results with respect to type of cultural participation are remarkably similar to Pääkkönen (1999), discussed in section 3.3.5.

The refusers less often participate in sports than the respondents. A distinction between types of sports makes no difference at all. For those who participate in sports, the share of team sports (football, volleyball, hockey, etc.), which might be an indicator of social participation, is identical in all groups. Refusers are also less often readers. This result is much less outspoken, and the position of the control group much less clear. Refusers and control groups both less often read books, which might suggest a design effect due to using a different questionnaire. Refusers less often buy books than the AVO respondents; the control groups even less. And finally, refusers are substantially less often member of a library than both refusers and the control group. All in all, one might conclude that the refusers are less involved in books and reading than the regular respondents.

Apart from involvement in classical culture, sports and reading, refusers also seem to be less active socially than the AVO respondents and the control group. Among the former, the percentage of those who are member of no organization at all is substantially higher than among the respondents. Table 9.8 gives more insight in what this might mean[2].

2 One reservation should be made here. The number of item missings on the yes/no membership questions is rather high in the AVO (about 6%) and almost nonexistent in the follow-up survey. In table 8.8 the scores that are missing have been recoded to non-membership for those respondents who scored yes on membership of at least one other organization. After this correction, approximately 1% of the respondents remained who had skipped the entire block. They are not included in table 8.8.

Table 9.7 Substantive variables respondents, followAup group, control group %

	VO response	followAup	control
health 1*** 2***			
very good	23.1	20.0	12.4
good	58.1	56.2	65.0
moderate	14.9	18.3	17.5
poor	1.9	4.7	5.1
very poor	2.0	0.9	0.0
(n) =	5981	235	177
gives care 1** 2* 3***			
no need	81.3	86.4	75.7
no care	9.6	10.2	14.7
care	9.1	3.4	9.6
(n) =	6125	235	177
cultural participation 1*** 3**			
both classical and popular	30.5	26.4	29.4
classical only	22.4	13.2	23.7
popular only	16.7	22.1	17.5
no participation	30.4	38.3	29.4
(n) =	6125	235	177
number of sports 1***			
0	39.8	50.2	42.4
1	23.5	26.8	26.0
2	17.2	11.1	16.9
3	9.7	5.5	4.5
4+	9.8	6.4	10.2
(n) =	6125	235	177
team sport			
no	86.1	86.3	86.3
yes	13.9	13.7	13.7
n active sports participants	3690	117	102
read book 1*** 2**			
more than 6 months ago	25.6	33.6	34.3
3-6 months ago	5.1	6.8	7.4
2-3 months ago	9.3	7.7	6.3
1 month ago or less	60.0	51.9	52.0
(n) =	6017	235	175
buy books own reading 1* 2***			
no	42.0	46.0	51.4
yes	58.0	54.0	48.6
(n) =	6018	235	177

Table 9.7 Substantive variables respondents, followUp group, control group % (cont.)

	VO response	followUp	control
library member 1*** 3**			
no	64.0	72.3	62.7
yes	35.5	27.7	37.3
(n) =	5989	235	177
membership associations 1***			
0	31.0	42.1	36.7
1	30.6	29.4	28.8
2	20.6	15.3	20.9
3	10.5	5.5	6.8
4+	7.3	7.7	6.8
(n) =	6125	235	177
arts and crafts and hobbies 1*** 2***			
0	85.9	63.8	59.9
1	11.7	24.9	23.2
2	2.0	8.2	7.9
3	0.4	1.3	4.5
4	0.0	0.0	0.6
5	0.0	0.0	0.6
6	0.0	0.4	1.1
7	0.0	1.3	2.3
(n) =	6125	233	177
sports programmes on tv/radio 1*** 2***			
once a week or more	43.8	57.4	55.4
1 to 3 times a month	12.5	11.1	13.6
less than once a month	8.2	9.8	10.2
no	35.5	21.3	20.9
(n) =	5978	235	177
art programmes on tv/radio 1*** 2*** 3*			
no	60.4	64.3	56.5
yes	39.6	35.7	43.5
(n) =	6079	235	177
watches films on tv 1*** 2***			
once a week or more	26.3	27.2	27.4
1 to 3 times a month	25.6	23.8	29.7
less than one a month	20.2	11.5	12.6
no	27.8	37.4	30.3
(n) =	5997	235	175
household owns vcr 1** 3**			
no	21.6	28.9	20.3
yes	78.4	71.1	79.7
(n) =	6086	235	177

Table 9.7 Substantive variables respondents, followAup group, control group % (cont.)

	VO response	followAup	control
pc use 1*** 2*** 3*			
unknown	2.5	0.0	0.0
no pc	41.6	55.7	42.4
use unknown	2.5	0.0	0.6
no use	11.8	6.4	10.2
user, off line	19.9	10.2	15.3
frequently on line	21.7	27.7	31.6
(n) =	6125	235	177

1 difference AVO-response and follow-up
2 difference AVO-response and control group
3 difference follow-up refusers and control group
Significance: ***0.01 level; **0.05 level; *0.10 level

Table 9.8 Membership organizations (%)

	regular avo	followAup survey
political party, society	6.2	7.7
labour union, employers' organization	18.9	17.0
social movement, pressure, lobby group (Amnesty International, third world shop)	9.8	9.8
protection of nature, environmental organization (Greenpeace)	28.0	20.0
educational association, school society	8.8	6.0
women's association	7.4	10.6
pports club	26.2	26.0
choral, musical, drama society	8.5	8.5
hobby society	8.2	8.9
youth centre or association, scouting	2.8	5.5
other society or association	18.9	14.5
N	6043	235

With respect to the membership of organizations respondents and refusers are remarkably similar. Respondents belong more often to nature conservancy organizations. They are more often a member of a parents' association or school society, which is mainly due to the fact that respondents with children are more often a member of this type of organization than refusers (15.2% as compare to 6.8%). The refusers are more often member of a women's organization, which is surprising because there are slightly more males among them. Also surprising is that the membership of sports clubs and choral, music and drama societies is almost identical, despite the differences mentioned above in cultural and sports participation. Refusers are more often members of youth associations. This is particularly true for the younger refusers. As the number of refusers younger than 35 in the

follow-up survey is small, however, it is hard to draw firm conclusions from these results. Finally, the percentage of members of 'other' organizations is higher among respondents than among refusers.

So far, the results seem to indicate a slightly lower social and cultural participation among refusers. This is in flagrant contradiction to their relatively high participation in arts and crafts and hobbies in table 9.7. The control group results, however, indicate that this might be due to the rephrasing and reducing of the questionnaire in the experimental study, as the control group too is much more active in these leisure activities. An inspection of the results of the different activities in table 9.9 suggests that there occurred a substantial design effect due to the shortening of the questionnaire. The regular questionnaire presented a complicated matrix in which one should score membership of organizations, the following of classes and so on, emphasizing that only serious activities were to be taken into account, and not just knitting and kin flicks. The reduced questionnaire for both the follow-up survey and the control group mentioned these restrictions in an explanatory note, which must have received much less attention from the respondent or the interviewer, considering the large differences in, especially, drawing, textural arts and photography. Because of this design effect, participation in arts and crafts will not be used in the comparison of respondents and refusers.

Table 9.9 Participation arts and crafts and hobbies (%)

	regular VO	followAup survey	control group
drawing	2.8	12.9	17.5
sculpting	1.2	4.7	8.5
textural arts	1.5	12.0	11.9
singing	5.3	6.9	13.0
music	2.8	6.4	10.2
drama	2.5	3.4	8.5
photography	1.0	10.3	10.7
N	6125	233	177

Table 9.7 comprises three types of TV watching, namely watching sports, watching art programmes and watching films. These variables are included because sports participation and cultural participation (classical and popular) seem to distinguish between respondents and *refusers*. TV watching patterns may show whether these differences are due to differences in outgoing behaviour or substantive interests. *Refusers* appear to be avid sports viewers, which denies their lack of interests in sports, but confirms the impression of this group as more passive than regular respondents, if not as inveterate couch potatoes. A breakdown according to gender in table 9.10 illustrates this even better. Almost 80% of the men in the follow-up

survey watch sports programmes at least once a week or more, compared to well over 60% in the regular AVO. For women this is 43% compared to 33%. The results from the control group, however, indicate that design changes may be the underlying cause of these differences. And indeed, although the questions are identical in this case, the context differs. The regular AVO was conducted from September 1999 to January 2000, the follow-up survey and the control group were conducted for the most part in June and July 2000. In these months the Roland Garros and Wimbledon tennis championship, the Tour de France and the European Championship Football were broadcasted. This championship, in which the Netherlands was knocked out in the semi-finals on the 29th of June (Italy won the penalties after a 0-0 draw) attracted so much public attention in the Netherlands that it is surprising that anyone answered 'Never' to the rather general questions 'Do you ever follow sports programmes on radio or TV?' and – if yes – 'How often do you do that on average?'. These outcomes show the influence of context, the fact that rather general questions may be influenced by temporary occurrences, and that these questions may not be very helpful in comparing respondents and refusers.

Table 9.10 Sports programmes on TV/radio (%)

	men			women		
	regular avo	follow-up survey	control group	regular avo	follow-up survey	control group
once a week or more	62	78	70	33	43	45
1 to 3 times a month	13	9	9	12	12	17
less than once a month	5	5	9	10	13	11
no	20	7	11	44	31	28

The results above cast doubt on the usability of two other questions on media use. The results indicate that refusers less frequently follow art programmes and more often never follow films on TV. These results differ substantially from those from the control group. The time-consuming sports viewing that went on during the follow-up survey will have influenced, however, the attendance to other TV programmes and this may also have biased the answers to the more general questions on media use. Therefore, TV attendance will not be used in the comparison of classes of nonparticipants.

Two substantive variables remain in table 9.7 that seem less liable to contextual effects and where the question texts had not been changed. Firstly, it turns out that the refusers less often possess a VCR, an appliance that was quite common in Dutch households at that time. VCRs are especially popular in families with children. Both among respondents and refusers with children the percentage of VCR owners is 92%. If there are no children in the family, the percentage VCR ownership is 68% for respondents and 61% for refusers.

Secondly, PC ownership is much less frequent among refuser households, although – if they own a PC – they use it more often to e-mail and browse the internet, an effect also mentioned by Väisänen (2002), and discussed in section 3.3.5 and 4.2.1. The bias in PC ownership is rather strong and has been extensively analysed by Schouten, Bethlehem and Stoop (2002). Among scores of variables this was the only effect that stood out strongly, although the effect can be expected to be smaller in future rounds of the AVO as PC ownership and use is increasing in the Netherlands. In a discussion at the International Workshop on Household Survey Nonresponse in Copenhagen in September 2002, it was suggested that the absence of PC use might be related to form illiteracy, which in turn might be a cause of unit nonresponse.

The results in this section suggest that the differences between respondents and refusers are small and that several differences (urbanicity, age, family composition) will at least partly be due to the fact that the target population of the refusers were those who did not respond on the regular survey. Therefore, the follow-up survey should to a certain extent complement the regular survey rather than be similar. Other differences may have been insignificant due to the small sample size of the follow-up survey. A number of differences with respect to the survey variables in table 9.7 will most likely have been caused by changes in the design and phrasing of the questions. This outcome illustrates the importance of the control group without which a number of effects would not have been identified as spurious. Finally, differences between refusers and respondents may have been a consequence of composition effects and be related to background variables rather than response behaviour. If this were the case, nonresponse in the AVO would be Missing at Random (MAR), which can be adjusted for by weighting for nonresponse, rather than the more serious case of Missing Not at Random (see section 2.3). The next section will combine background variables and survey variables in a single model to find out the extent of bias in the AVO.

9.3.2 *Classes of nonparticipants*

Table 9.11 summarizes the relationship between socio-environmental, background and survey variables in a multinomial regression of a classes of nonparticipants model (see section 5.2). Refusers are compared with reluctant respondents, who refused at least once, hard to contact respondents, who did not refuse but needed more than two calls, and easy respondents (the reference group, see annex 7.1) who were not particularly hard to reach nor particularly reluctant to cooperate. The results for the hard-to-contact group are largely in line with the findings presented in chapter 7, and those for reluctant respondents to those in chapter 8. The small group of refusers does not significantly differ on any of the background variables from the reference group, the easy-to-contact and easy-to-persuade respondents, except for the fact that they more often have a listed telephone number. They do differ in terms of cultural participation and sports activities. Whereas among reluctant respondents the category 'classical culture only' is relatively overrepresented,

this category is clearly underrepresented among refusers. As opposed to all respondents, sports activities are rather unpopular among refusers. Refusers also indicate more often that they do not belong to a religious group. The largest difference between refusers and respondents turns out to be PC ownership and use, a factor that had not been expected to be of importance beforehand, but which proved to have a very pronounced effect (see above). Among scores of variables this was the only effect that stood out strongly. It may not be repeated in the future as PC penetration is rapidly increasing in the Netherlands.

This analysis and the tables in the previous section show that refusers differ from respondents with regard to a small number of survey variables, that these differences are small and are generally not significant when differences in sample composition are taken into account. This conclusion is illustrated by figure 9.2, which presents the outcome of a categorical principal components analysis (CATPCA, see annex 9.1) of a wide range of background and survey variables and classes of nonparticipation. The first principal component, presented on the horizontal axis, distinguishes between young (right) and old persons (left), the second between those who live in rural (below) and urban areas (above). Age follows a curved pattern, reflecting the diverging position of young persons. Income and education follow the same left-right pattern as age. In the right upper quadrant are the categories reflecting active social and cultural participation (religious, avid reader, cultural participation, internet use, and sports). In the left lower quadrant lie the opposite categories.

Response class plays a minor role in this analysis. The four response classes are close to the centre of the figure, which illustrates that response behaviour is largely independent of background and survey variables. The hard to contacts are to be found in between the big city dwellers and active participation, refusers on the other hand share some big city characteristics but are located among respondents who do not participate socially and culturally. Reluctant respondents are exactly in the centre of the plot. This is probably due to the fact that they do not quite fit in the picture, as we by now know that among them are many big city dwellers and few male singles, whereas these categories in real life belong together. Finally, easy respondents seem to live slightly more often in less urban environments.

Table 9.11 Net sample (respondents), classes of nonparticipants, multinomial regression (n 6360)[a]

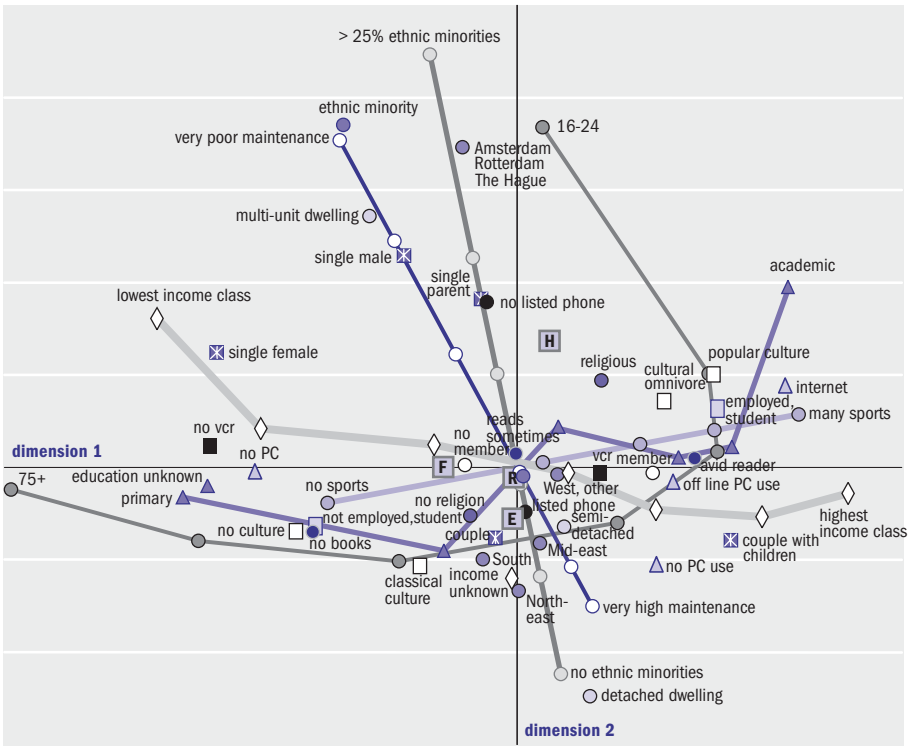
	β			exp β			standard error		
	re-fuser	reluc-tant	hard to contact	re-fuser	reluc-tant	hard to contact	re-fuser	reluc-tant	hard to contact
intercept	–2.419	–0.039	–0.380	0.089	0.962	0.684	0.450	0.212	0.222
age (years)	–0.008	–0.002	–0.009	0.992	0.998	0.992	0.006	0.003	0.003
age: absolute deviation mean (years)	–0.008	–0.012	–0.008	0.992	0.988	0.992	0.008	0.004	0.004
listed phone number	0.504	–0.101	–0.257	1.655	0.904	0.774	0.200	0.082	0.081
Amsterdam, Rotterdam, The Hague	–0.216	0.688	0.633	0.806	1.990	1.883	0.225	0.090	0.089
good maintenance	–0.210	–0.259	–0.259	0.810	0.772	0.772	0.166	0.077	0.080
detached dwelling	0.002	–0.422	–0.280	1.002	0.656	0.756	0.202	0.104	0.108
male	–0.162	–0.765	–0.052	0.850	0.465	0.949	0.145	0.076	0.070
single	0.138	–0.374	0.294	1.149	0.688	1.342	0.175	0.093	0.085
child at home	–0.298	–0.109	–0.366	0.742	0.897	0.694	0.180	0.083	0.087
employed, student	0.172	–0.135	0.314	1.188	0.874	1.369	0.186	0.082	0.089
higher education	–0.152	–0.508	–0.031	0.859	0.602	0.970	0.200	0.099	0.088
cultural participation (df=9)									
both classical and popular	–0.206	–0.030	0.130	0.814	0.970	1.139	0.203	0.095	0.098
classical	–0.563	0.228	0.168	0.569	1.256	1.183	0.219	0.090	0.098
popular	0.166	–0.075	0.150	1.181	0.928	1.162	0.212	0.109	0.110
no sports activities	0.409	–0.030	0.000	1.506	0.971	1.000	0.153	0.072	0.075
pc in household	–0.875	0.178	0.003	0.417	1.195	1.003	0.208	0.085	0.088
use of internet, e-mail	0.944	0.068	0.138	2.571	1.070	1.148	0.215	0.093	0.092
no religious affiliation	0.378	–0.111	0.003	1.459	0.895	1.004	0.143	0.071	0.070
		Wald			**sign**		X2	**df 3**	**sign.**
intercept	28.90	0.03	2.92	0.000	0.854	0.088			
age (years)	1.61	0.40	8.46	0.205	0.526	0.004	9.30		0.026
age: absolute deviation mean (years)	1.02	8.97	4.15	0.313	0.003	0.042	10.65		0.014
listed phone number	6.35	1.49	10.16	0.012	0.222	0.001	20.23		0.000
Amsterdam, Rotterdam, The Hague	0.92	58.66	51.14	0.338	0.000	0.000	88.00		0.000
good maintenance	1.61	11.22	10.40	0.204	0.001	0.001	17.61		0.001
detached dwelling	0.00	16.58	6.70	0.993	0.000	0.010	20.56		0.000
male	1.25	101.33	0.56	0.264	0.000	0.455	112.17		0.000
single	0.63	16.16	11.90	0.428	0.000	0.001	41.43		0.000
child at home	2.76	1.73	17.85	0.097	0.188	0.000	19.08		0.000
employed, student	0.85	2.70	12.52	0.355	0.100	0.000	21.65		0.000
higher education	0.58	26.04	0.12	0.448	0.000	0.729	28.92		0.000
cultural participation (df=9)							27.08		0.001
both classical and popular	1.02	0.10	1.76	0.312	0.751	0.185			
classical	6.63	6.38	2.91	0.010	0.012	0.088			
popular	0.62	0.47	1.86	0.432	0.493	0.173			
no sports activities	7.19	0.17	0.00	0.007	0.681	0.999	7.80		0.050
pc in household	17.60	4.36	0.00	0.000	0.037	0.972	26.06		0.000
use of internet, e-mail	19.23	0.53	2.26	0.000	0.469	0.133	20.79		0.000
no religious affiliation	7.01	2.46	0.00	0.008	0.117	0.960	10.78		0.013

a Contrast with easy-easy respondent; Chi-square 693 (df=54); Cox and Snell R^2 0.103 Nagelkerke R^2 0.115.
Reference categories (=0) Other regions, no cultural participation, not employed, student.

Refusers in this analysis are most similar to the respondents that are socially and culturally inactive. This might be even more outspoken if information had been available on those who refused to cooperate in the follow-up survey. From table 9.1 it is known that these double refusers are older than all other groups, and from figure 9.2 that age and inactivity are related.

From the previous section one might conclude that refusals show a lower social participation than respondents. In table 9.11 only a few of these characteristics remained that were significant in a model that also included background variables. Figure 9.2 appears to support the conclusion that nonresponse bias in the AVO1999 was very small, but also gives a somewhat impressionistic view of the widespread lower participation of refusals. The effects are small, but consistent. In addition, there is no evidence at all that temporary refusals are more similar to final refusals than other survey participants.

Figure 9.2 Categorical PCA, classes of nonparticipants



Variance explained dimension 1: 20%, dimension 2 15%;
Cronbach's alpha dimension 1: 0.78, dimension 2: 0.68.

9.4 *Bias and robustness: an analysis of sensitivity*

In the previous sections differences were found between refusals and respondents. This section will focus on the size of the effects and give an estimate of the amount of bias. This is done firstly by a sensitivity analysis that investigates worst-case scenarios and secondly by a randomization study that investigates whether the differences could have been caused by sampling errors.

In table 9.12 three variables are presented that distinguish between respondents and refusers. The first row gives the distribution in the regular AVO across the different categories of the variable. The second row indicates what the percentage would have been in the unlikely case that none of the nonrespondents would have belonged to a particular category, and the third row the equally unlikely case that all nonrespondents would have fallen in this category. For instance, the cell at the intersection of the second row and second column of 'Sports' indicates that if no nonrespondents would participate in any sports activities the percentage of the total sample that would participate in at least one sport would be 43% (compared to 60% measured in the AVO). If all nonrespondents would do sports, the percentage of sports participants would be 72%. The fourth row presents the percentages observed in the follow-up survey, and the fifth the percentages for the total sample, assuming that the *refusers* who participated in the follow-up survey represent all nonrespondents. The sixth and the seventh row do the same, but give lower and upper limits to correct for the small sample size of the follow-up survey, based on 95% confidence limits of the outcome of the follow-up survey. And finally, the eighth and ninth row give the percentages assuming that the total contacted sample of the follow-up survey represents all nonrespondents, and that of those who refuse to participate in the follow-up survey either none (8th row) or all (9th row) belong to a particular category.

Table 9.12 stimates of bias due to nonresponse

cultural participation	cultural omnivore	classical only	popular only	no cultural participation
AVO	31	22	17	30
no nonrespondents in category	22	16	12	22
all respondents in category	51	45	41	51
follow-up survey	26	13	22	38
refusers represent all nonrespondents	29	20	18	33
refusers -95% sample margin	28	19	17	31
refusers+95% sample margin	32	23	19	33
no refusers follow-up survey in category	27	19	16	30
all refusers follow-up survey in category	35	27	25	38

sports	yes	no
AVO	60	40
no nonrespondents in category	43	57
all respondents in category	72	28
follow-up survey	50	50
refusers represent all refusers	57	43
refusers -95% sample margin	56	41
refusers+95% sample margin	59	44
no refusers follow-up survey in category	53	39
all refusers follow-up survey in category	61	47

pc	no pc	no use	use, no internet or eAmail	use, internet and/or eAmail
AVO	43	15	20	22
no nonrespondents in category	30	10	14	16
all respondents in category	60	40	44	45
follow-up survey	56	6	10	28
refusers represent all refusers	47	12	18	24
refusers -95% sample margin	46	12	17	23
refusers+95% sample margin	49	13	19	26
no refusers follow-up survey in category	42	12	17	22
all refusers follow-up survey in category	50	20	25	30

Based on these theoretical upper and lower limits, based on the entire eligible, contacted unweighted sample, it seems that cultural participation in the AVO is slightly overestimated, or 'no participation' is underestimated. The unweighted outcome of 'no cultural participation' in the regular survey is 30%, whereas the outcomes of the follow-up survey show that the real percentage should be in between 31% and 33% (95% sample margin) or 30% and 38% (dependent on the assumptions concerning

the nonrespondents on the follow-up survey. It is also likely that the AVO slightly overestimates sports participation and the possession of a PC (but might underestimate internet use). It should be noted that these are the variables that showed the strongest effects of all variables in the follow-up survey and that part of these effects will be related to differences in the composition of the different groups, and mitigated by post-stratification weighting. The next analysis takes these differences in composition into account.

Table 9.13 gives another impression of the significance of the analysis outcomes. The multinomial regression in table 9.11 was repeated 1,000 times. In each analysis all 235 *refusers* were used and each time a random subsample of 235 easy respondents, 235 hard to contact respondents and 235 initially reluctant respondents. This removes the effects of sample size and treats each group in a similar way as the respondents in the follow-up sample, namely as a small subgroup of the larger, unobserved group. Table 9.13 presents statistics on the significance of the chi-square that reflects the loss of likelihood if a particular variable was removed from the complete model. A significance of 0.01 means that a variable is very important in the regression, of 0.05 and 0.10 somewhat less significant and larger values indicate that a variable might as well be removed from the analysis. In table 9.11 this significance is given under 'total' for all respondents; in this original analysis of a much larger sample all variables contributed significantly to the model.

Table 9.13 Results randomization study multinomial regression (1000) chi-squares

	max	min	mean	median	N<.10	N<.05	N<.01
age (years)	1.000	0.001	0.423	0.391	153	93	17
age: absolute deviation mean (years)	1.000	0.000	0.461	0.444	142	80	21
listed phone number	0.391	0.000	0.025	0.009	938	871	521
Amsterdam, Rotterdam, The Hague	0.112	0.000	0.003	0.000	999	990	923
good maintenance	1.000	0.000	0.423	0.379	179	113	30
detached dwelling	0.995	0.000	0.332	0.280	262	163	55
male	0.565	0.000	0.015	0.001	963	931	788
single	0.955	0.000	0.105	0.035	689	565	309
child at home	0.991	0.000	0.326	0.272	289	178	49
employed, student	0.994	0.000	0.252	0.151	396	262	92
higher education	0.999	0.000	0.278	0.188	370	251	90
cultural participation	0.786	0.000	0.081	0.048	728	512	139
no sports activities	0.672	0.000	0.068	0.033	777	600	227
pc in household	0.043	0.000	0.001	0.000	1000	1000	982
use of internet, e-mail	0.110	0.000	0.003	0.001	999	997	928
no religious affiliation	0.648	0.000	0.090	0.052	687	481	147

Table 9.13 shows that in the much smaller 1,000 samples only a small number of variables remain (frequently) significant. The first column shows that in 1,000 subsamples PC ownership always remains below the 0.05 level (the maximum chi square in 1,000 subsamples is 0.043) and is thus always discriminating between the classes of nonparticipants. The mean value for this variable is .001 and the median .000. In 982 of the 1000 analysis the significance level is lower than 0.01. Other variables that remain significant in most of the cases are urbanicity (Amsterdam, Rotterdam and The Hague), sex (male), and the use of internet. Significant at a 0.10 level in more than two thirds of the cases are family composition (single), cultural participation, sports activities and religious affiliation.

9.5 *Conclusion*

What can be said after all these tables and figures and analyses in this and the previous chapters? This study started from a concern about nonresponse bias in the AVO1999. This concern appears to have been largely unwarranted. There may be a small bias, but nothing to be really concerned about, according to the previous sections. Does this mean that nonresponse is not to be worried about? Not at all, because it cannot be assumed that the present results hold for other surveys at other times in other countries and on other topics. What this study has shown is how nonresponse bias can be studied. The combination of detailed call records, standard outcome codes, interviewer' observations of the neighbourhood and the dwelling and a follow-up survey among refusals can provide the required information on persistent refusals and on the extent to which they differ from different types of respondents.

The drawback of this approach is that call records are hard to keep in face-to-face surveys and that a follow-up survey among refusals is very expensive and difficult to carry out. In the previous sections many minor imperfections and questions about the completeness and quality of the contacts forms and the experiment have been raised. It is beyond doubt that the approach of the present follow-up survey can be approved upon. But, and that is most important, it can be done and opens up the black box of nonresponse.

Besides giving an indication on the size of the nonresponse bias, the combined approach of follow-up survey and very detailed information on the fieldwork made it possible to compare temporary, converted refusals – also called reluctant respondent – with final refusals. In the present study converted refusals were by no means more similar to final refusals than cooperative respondents. As a result, there is no conclusive evidence to uphold the model of nonparticipants, nor that enhancing cooperation through refusal conversion improves survey accuracy . These results definitely cannot be generalized to other situations and other surveys. They prove, however, that one should not assume – without additional evidence – that temporary

refusals can be used as a proxy for final refusals, nor that higher response rates – by converting temporary refusals – will improve survey quality and lead to more accurate survey estimates.

The results of the present study can be seen as a plea for not indiscriminatingly trying to enhance response rates, but for settling for slightly lower ones and use the remaining funds for a direct study of nonresponse bias. This strategy will be discussed in chapter 11. It has not been followed in the European Social Survey (ESS) where a uniform target response rate of 70% was set in all participating countries. In this survey no follow-up survey among refusals was conducted, but detailed contact forms were kept and keyed that were very similar to the ones in the AVO1999. Because of similarities in paradata collection, some results of the ESS nonresponse work will be presented in chapter 10, in order to provide a more general view on nonresponse analysis than could be provided by the AVO alone.

10 Nonresponse and the European Social Survey[1]

10.1 Comparing and analysing response processes and response rates

The recurrent message of the previous chapters was that the composition of the nonresponse and the bias due to nonresponse are more important than the response rates as such. This implies that comparing response rates over time or between surveys or across countries may just scratch the surface of possible differences. In a comparison of response rates over time, for instance, one should take into account that the ratio of noncontacts to refusers may have changed, or that differences in response rates may have been inflated or deflated by changes in fieldwork procedures (length of fieldwork period, number of calls, use of incentives) and fieldwork costs. Even response rates that are stable over time can disguise differences in response composition and different types of response bias. Detailed information on fieldwork and response behaviour, as presented in the previous chapter, can uncover these differences.

This is even more pertinent in cross-national and cross-cultural survey research, as different countries have may distinct survey climates and survey traditions, and ease of contact and reluctance to cooperate are likely to differ among cultures. Johnson et al. (2002, p. 68) suggest that social participation patterns, socio-economic opportunities and communication styles may influence cultural values and survey nonresponse. De Leeuw and De Heer (2002, pp. 52-53) in their comparative analysis of nonresponse in the Labour Force Survey found differences in noncontact rates to be associated with differences between countries in average household size and percentage of young children, and differences in refusal rates with economic indicators (percentage of unemployed, inflation rate). Couper and De Leeuw (2003) compiled evidence that countries differ not only in response rates but also in the composition of the nonresponse due to differences in design and fieldwork procedures. They draw attention to the paucity of strict comparative cross-cultural and cross-national research, and end with a number of recommendations, among which the need to study the reasons for and impact of different response rates in comparative surveys, the differences in response composition and the differences between respondents and nonrespondents across countries.

1 This chapter is partly based on data quality papers and congress papers from Jaak Billiet, Michel Philippens and Geert Loosveldt from Leuven University, the institute responsible for data quality in the ESS. The author cooperated with the Leuven group in developing the ESS contact forms and in some of the analyses of nonresponse. Information on all aspects of the survey can be found on www.europeansocialsurvey.org

Ideally, one should try to reduce variation in response rate and in the components of nonresponse. At a minimum, they strongly recommend reporting components of nonresponse and details of fieldwork and sample realization. In addition, they emphasize that ... *Nonresponse reduction requires careful planning, the collection of auxiliary data to inform decisions about nonresponse follow-up, and the use of such information to enhance adequate postsurvey adjustment for nonresponse. In cross-cultural research, this implies much more careful consideration of the likely impact of nonresponse on key statistical estimates than simply maximizing or reporting response rates in each country or subgroup* (p. 177).

The recommended measures had already been implemented in the European Social Survey 2002. Section 10.2 will introduce this comparative study, describe the specifications for and organization of fieldwork, and show how high survey standards were pursued. Section 10.3 sets out the design and implementation of contact forms and the paradata (see section 2.4.3) outcomes. Basic results for contactability and reluctance to cooperate will be presented in section 10.4 and 10.5. The final section looks back on the feasibility of the instrument and on possible nonresponse bias uncovered by the information from the contact forms. This chapter will focus on approaches to study nonresponse and the collection of paradata, i.e. on the instrument rather than on specific outcomes. The results of the first analyses of the ESS 2002/ 2003 nonresponse data by Michel Philippens, Geert Loosveldt and Jaak Billiet from Leuven University have already been used by national teams in order to improve fieldwork and enhance response rates in round 2, fielded in Autumn 2004.

10.2 *European Social Survey*

The European Social Survey is a bi-annual face-to-face survey of values, attitudes, beliefs and behavioural patterns in the context of a changing Europe. Its major aim is to provide data to social scientist for both substantive and methodological studies and analyses. Data from the first round of the ESS were collected in 2002/2003 in 22 European countries and are available from the ESS data archive, the Norwegian Social Science Data Service (NSD). It is funded under the EU 5th and 6th Framework Programme, and receives additional funding from the European Science Foundation. Within each country, the fieldwork is (mainly) sponsored by the national science organization.

The substantive aims of the ESS have been pursued by involving a large number of academic experts in the construction of the draft core questionnaire which comprises questions on occupation and social structure, religious affiliation and identity, ethnic and national identity, political trust, party affiliation and voting behaviour, media consumption, value orientations and social exclusion. In addition to the 30 minute core, each round comprises two or three more topic-specific modules to measure a variety of academic and policy concerns and debates that require examination in depth. The topics have to be appropriate to academic and policy

needs within Europe, but also its suitable for multicultural, multinational measurement. The subject matter of the topic modules is determined after a call for proposals from small multinational teams of European scholars.

The methodological aim of the ESS is to pioneer and 'prove' a standard of methodology for cross-national attitude surveys that only the best national studies usually aspire to. This manifests itself in the pursuit of optimal comparability across countries or in the prevention of unnecessary variation (De Heer, 2000). According to De Heer (1999, pp. 136-137) different factors may explain cross-national differences in survey quality and survey outcomes:

- General design factors, such as mode of data collection, panel versus cross-section, person or household, proxy allowed, substitution allowed;
- Practical fieldwork strategies, such as formal call scheduling, number of contact attempts, use of refusal conversion, interviewer incentive, respondent incentive;
- Factors related to survey organization, such as voluntary participation in surveys, employment condition of interviewers.

Variations due to the first factor were eliminated in the ESS by centrally specifying a single survey design for all participating countries. The guidelines and specifications for the first two rounds of the survey are available at the ESS website (see also Annex 10.1). A target response rate of 70% and a target noncontact rate of 3% were specified in order to minimize the chance that cross-cultural differences would mainly reflect differences in response rates. As mentioned before, even similar response rates may hide a different composition of nonresponse and different types of bias, and a response rate of 70% does not necessarily leads to lower nonresponse bias than a lower one. Therefore, the target response rate was accompanied by process requirements to ensure that noncontacts were minimized and temporary refusals would be converted. Therefore, variations due to the second factor were minimized partly by the central fieldwork specifications, comprising a number of process requirements such as a minimum of four calls, and partly by recommendations and suggestions for fieldwork and response enhancement.

Variations due to the third factor remain as the fieldwork was carried out by national organizations. Publicly available national technical summaries show how the survey was implemented in each participating country and give information on national particularities within (and occasionally outside) the specifications such as interviewer assignments, interviewer training, interviewer remuneration, the use of incentives, etc. etc. In addition to providing guidelines and specifications, national teams received extensive support from expert panels in sampling and translation, and sophisticated studies were carried out on reliability and validity of survey questions before and after fieldwork. And finally, not necessarily to minimize variation but to enhance transparency, standard procedures were specified for collecting paradata and documenting the survey. The paradata cover all aspects of the fieldwork

process, including interviewer identifications, respondents selection, call records, standard call and final outcomes, reasons for ineligibility, refusal and not being able to cooperate, and interviewer' observations of neighbourhood and dwelling. The paradata outcomes have already been used to check whether the fieldwork has been carried out according to the specifications, to compare nonresponse processes, composition and bias across countries, and to improve fieldwork in future rounds.

The wealth of information on the fieldwork process and response behaviour, and on the deviations that occurred, is publicly available on the ESS website, which makes the ESS not necessarily the best survey ever, but certainly the most transparent cross-cultural survey ever.

10.3 Paradata

10.3.1 Design and implementation of contact forms

For the first round of the ESS standardized contact forms were envisaged. Stoop et al. (2003) describe the development of these forms.[2] From discussions with non-response researchers, it turned out that in many survey organizations the results of contact attempts and reasons for refusal were jotted down unsystematically for fieldwork logistics, to give interviewers or their supervisors information on how to handle individual cases. This information was generally not keyed, however, apart from final outcomes. There are, however, several organizations that routinely use standardized contact forms. Several members of the International Workshop on Household Survey Nonresponse, kindly forwarded their contact forms for an inventory, as a basis for the ESS contact form.[3] The resulting ESS contact form specified which information should be kept and keyed on the call pattern, the outcomes of calls, reasons for refusal, dwelling and neighbourhood characteristics, etc.

2 Silke Devacht, Leuven University, did most of the drafting of the contact forms. The Round 1 contact form for the address sample is available at http://naticento2.uuhost.uk.uu.net//fieldwork/contact_form_address.doc. The contact form specification in the ESS prescribed the information that needed to be collected during the fieldwork process. The implementation of the contact forms may have varied per participating country.

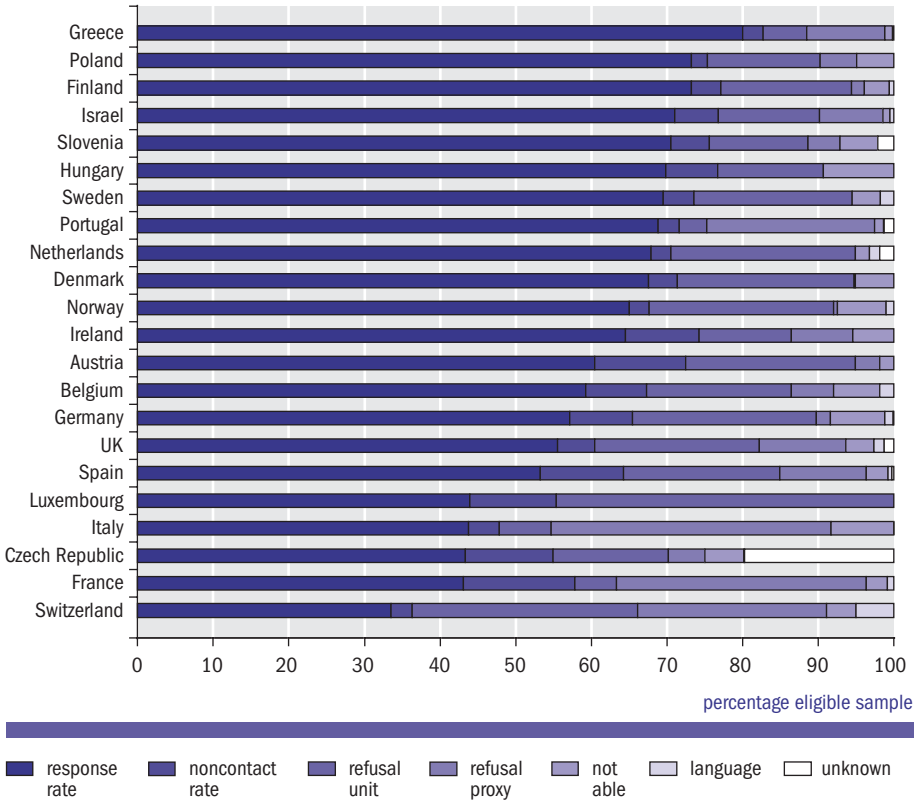
3 US National Health Interview Survey (Trena Ezzatti-Rice), Center for Disease Control/NCHS (Franklin Winters), University of Michigan, Ann Arbor (Mick Couper), General Household Survey, ONS (Dave Elliot), ONS (Mark McConaghy), British Household Panel Survey, NatCen (Peter Lynn), Danish Household Budget Survey, Statistics Denmark (Bo Møller), Statistics Finland (Paavo Väisänen), Leefsituatie Onderzoek Vlaamse Ouderen, CBSG (Marc Callens), Departement Sociologie, Leuven (Jaak Billiet, Geert Loosveldt, Ann Carton), Aanvullend Voorzieningengebruikonderzoek, GfK Nederland/SCP.

With the resulting paradata the fieldwork process can be monitored and supervised, it can be checked whether the fieldwork has been carried out according to the specifications, nonresponse rates, contact rates and cooperation rates can be computed, the bias due to nonresponse can be estimated, and recommendations for improving fieldwork in future rounds in subsequent rounds can be given.

By using standardized contact forms, it was assumed, data collection would be closer to a transparent and controlled process. In designing and using the contact forms, several problems arose. Firstly, although the ESS is based on strictly random sampling, sampling frames differed across countries. Therefore, versions had to be developed for address, household and individual samples. The first included a section that gave guidelines for the selection of households at multiple-household addresses and how to record the outcomes. The first two gave guidelines on how to select persons within households and how to record these outcomes. Secondly, there were differences in fieldwork procedures across countries. In some countries, where sampling was based on a list of named individuals, first calls could be made by telephone. In these countries the recording of call data was very cumbersome because the number of telephone contact attempts is generally much higher than the number of in-person contact attempts. The problems took two forms. Where calls were made centrally, recording was no problem at all, but the prescribed data format was, and where calls were made by the interviewer, the manual recording of each telephone call was prohibitive.

As not every fieldwork organization was familiar with keeping close track of the fieldwork according to the specified format and with using this information in fieldwork monitoring, the burden of recording and keying call information was considered to be rather heavy, and interviewers did not appreciate the extra work. Furthermore, the construction of the final response file required extra efforts and there was no clear relationship between the call records and the final outcome code. It also turned out that the contact forms were not complete or entirely clear, as codes had to be added during the fieldwork process (fraudulent interview) and other codes were difficult to fill in (is this a re-issue?). A number of countries would not deliver the observational data to be provided by the interviewer, as national confidentiality laws forbade this as an infringement of privacy, and some countries would not deliver reasons for nonresponse, for the same reason. Other countries provided contact forms for respondents only, or delivered very incomplete data.

Figure 10.1 Response overview ESS round 1 2002/2003



Source: National Technical Summaries (NSD, 2004)

10.3.2 *Quality contact form data and cross-national differences*

Compared to the surveys discussed by Couper and De Leeuw (2003), the variations in the ESS design are small, allowing a more accurate comparison of cross-cultural differences in response. Table 10.1 present an overview of the response rates in the European Social Survey. Figure 10.1 gives a visual overview of the final outcomes per country, ordered by response rates. The data come from the National Technical Summaries, drafted by the national coordinators and deposited at the NSD. The results show that not every country achieved a 70% response rate, although many did, even countries as the Netherlands for which this was generally considered to be next to unachievable. Despite the central specifications and the standard fieldwork procedures, there are vast differences in response rates, from 33.5% in Switzerland to 80% in Greece. The aim of setting a target response rate of 70% in order to minimize variation between countries has certainly not been achieved.

Table 10.1 Response overview uropean Social Survey 2002/2003

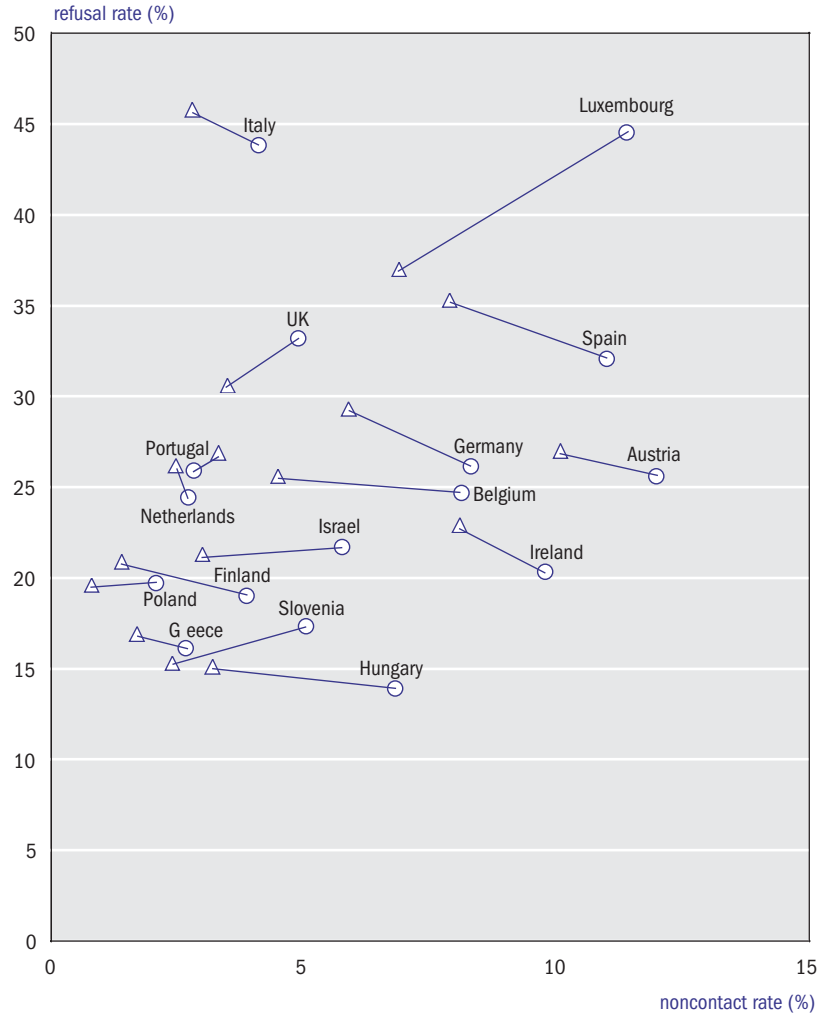
country	eligible sample units	resA ponse rate	nonA contact rate	refusal rate		incapacity rate[a]		unA known[c]	net sample
				total	by proxy[b]	total	language		
Austria	3736	60.41	12.04	25.67	3.27	1.87	0.00	0.00	2257
Belgium	3207	59.21	8.11	24.70	5.61	7.98	1.87	0.00	1899
Czech Republic	3139	43.33	11.63	20.01	4.81	5.32	0.13	19.72	1360
Denmark	2229	67.56	3.77	23.60	0.18	5.07	0.00	0.00	1506
Finland	2732	73.21	3.88	19.03	1.76	3.88	0.62	0.00	2000
France	3488	43.09	14.68	38.53	33.03	3.70	0.86	0.00	1503
Germany	5242	57.13	8.34	26.12	1.87	8.28	1.11	0.13	2919
Greece	3208	79.99	2.74	16.08	10.35	1.18	0.22	0.00	2566
Hungary	2412	69.86	6.84	13.93	0.00	9.37	0.00	0.00	1685
Ireland	3174	64.46	9.80	20.26	8.13	5.48	0.00	0.00	2046
Israel	3520	70.99	5.77	21.76	8.38	1.48	0.51	0.00	2499
Italy	2761	43.72	4.13	43.86	37.02	8.29	0.04	0.00	1207
Luxemburg	3535	43.93	11.46	44.61	0.00	0.00	0.00	0.00	1552
Netherlands	3484	67.85	2.70	24.37	0.00	3.16	1.32	1.92	2364
Norway	3132	65.01	2.65	24.87	0.51	7.47	1.05	0.00	2036
Poland	2881	73.24	2.08	19.75	4.82	4.93	0.00	0.00	2110
Portugal	2196	68.81	2.82	25.82	22.22	1.23	0.09	1.32	1511
Slovenia	2154	70.52	5.06	17.27	4.22	4.97	0.00	2.18	1519
Spain	3249	53.22	11.05	32.10	11.45	3.32	0.49	0.31	1729
Sweden	2878	69.46	4.07	20.95	0.00	5.52	1.84	0.00	1999
Switzerland	6097	33.48	2.80	54.81	24.95	8.91	5.04	0.00	2040
UK	3696	55.52	4.92	33.17	11.39	5.09	1.33	1.30	2052

a Respondent mentally or physically unable to cooperate throughout the fieldwork period; language barrier; respondent unavailable throughout the fieldwork period for other reasons.
b Refusal to cooperate by other person within the household or at the address, possibly within knowing whether the refuser is the target sample person.
c Percentage of sample units not accounted for.

Source: NSD (2004)

Table 10.1 also shows very large differences in noncontact rates across countries, which range from 2% in Poland to almost 15% in France. The refusal rate varies between 16% in Greece and 55% in Switzerland, and the percentage that is not able to cooperate between almost 0% in Luxemburg and 9% in Switzerland. The latter result indicates that part of these differences may be due to differences in reporting rather than differences in fieldwork outcomes. Closer inspection shows that in Luxemburg probably only a subset of the outcome codes has been used. Other indicators of difficulties with registration rather than with fieldwork are that in the Czech republic almost 20% of the sample units has not been accounted for, that Switzerland has a surprisingly large number of language problems (5% while the questionnaire was fielded in German, French and Italian) and that there are sizeable differences in the percentage of refusal by proxy as a share of the total number of refusals.

Figure 10.2 Noncontact rate and refusal rate ESS round 1 2002/2003 according to field report and contact form



△ Field report: National Technical Summaries (NSD, 2004)
○ Contact forms: Billiet and Philippens (2004)

Figure 10.2 compares the field outcome codes as given in the National Technical Summaries (NSD, 2004) and as computed from the call records by the Leuven team. This figure shows the large difference in noncontact rates and refusal rates between the participating countries. It also shows the differences between outcome codes computed by field organizations and calculated from the call records. The total response rate is generally very similar for both computation, but causes of nonresponse differ. In the figure 16 countries for which both types of information are

available are ordered according to response rates. In a large number of countries there is a small, identical shift between noncontact and refusal (Italy, Spain, Germany, Austria, Ireland, Finland, Greece). According to the call records the refusal rate is higher than reported by the national coordinators, and the noncontact rate lower. The most likely cause is that the outcome of the final call was a noncontact, but this final call was preceded by a refusal. In the analyses of the call records this will be scored as a refusal (Billiet and Philippens, 2004). In the UK, the call records indicate a smaller rate of both refusals and noncontacts. This means that according to the call records more nonrespondents were not able to participate. This is to an even higher extent true for Luxembourg, where according to the national technical summaries nobody was not able to participate. This can be redressed using the information from the call records.

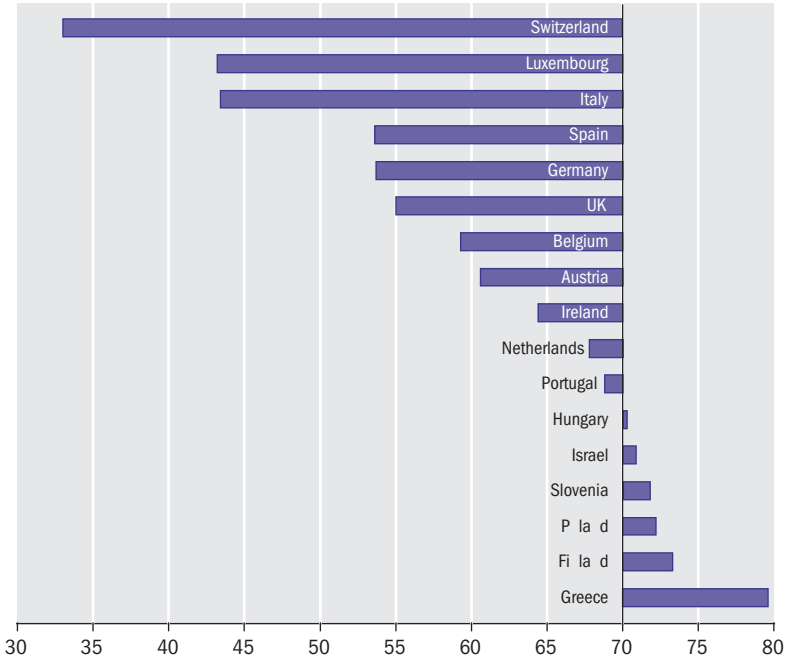
Annex 10.2 presents information from the National Technical Summaries on the use of incentives and refusal conversion strategies, as background to response rates. It shows that in several countries national coordinators mention general persuasion strategies under refusal conversion strategies, whereas others explicitly refer to re-approaching initially reluctant refusers. Other reporting differences have also occurred. Loosveldt et al. (2003) mention that there are large differences in the reasons for refusal that are reported that will at least partly be due to lack of standardized registration of reasons for refusal in different countries.

The outcomes so far give a first indication of the usefulness of keeping and keying call records and using standard contact forms. Without these, differences in noncontact and cooperation rates would have occurred that could partly be chalked up to different ways of computing standard outcomes. As the call records make it possible to compute standard outcome codes, these will be used in the remainder of this chapter, although only 16 of the 22 countries provided complete call records.

The message of this first look at the paradata is mixed. The bad thing is that, even in a centrally controlled, high quality survey as the ESS there are large differences in fieldwork registration, procedures and outcomes. The good thing is that the ESS is one of the view cases in which these differences have come to light and where the information on the fieldwork in previous rounds can and is used to improve fieldwork.[4] The subsequent sections will highlight how information from call records can give further insight in the fieldwork process, and in types of possible nonresponse bias and differences in nonresponse bias across countries, and can give suggestions for fieldwork improvements. As a background for this unravelling of the response process, figure 10.3 again presents the response rates on the ESS, now only for those countries with complete fieldwork records and as deviations from the target response rate.

4 Response rates in Switzerland were substantially higher in Round 2 than in Round 1.

Figure 10.3 Response rate ESS round 1 2002/2003 (deviations from target response rate 70%)



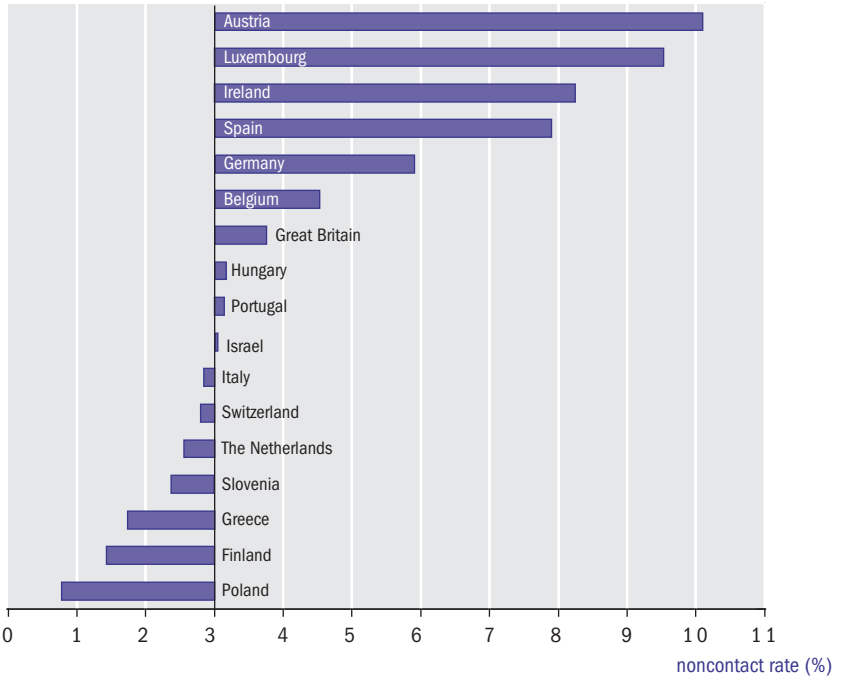
1. Countries with no or incomplete call records excluded
2. Data for Switzerland based on subsample with telephone recruitment

Source: Billet and Philippens (2004)

10.4 *Contacting respondents in the* ESS

Figure 10.4 depicts the noncontact rates calculated from the call records and shows that the specified maximum noncontact rate of 3% was not reached everywhere. From figure 10.3 and 10.4 it can be observed that countries with similar response rates, for instance Portugal and Poland, have fairly dissimilar noncontact rates. In addition, countries with a quite acceptable response rate (Ireland) can have a sizeable noncontact rate, which shows that by making more calls the response rate in Ireland could substantially increase.

Figure 10.4 Noncontact rate ESS round 1 2002/2003 (deviations from target noncontact rate of 3%)



1. Countries with no or incomplete call records excluded
2. Data for Switzerland based on subsample with telephone recruitment
3. Telephone recruitment allowed in Finland

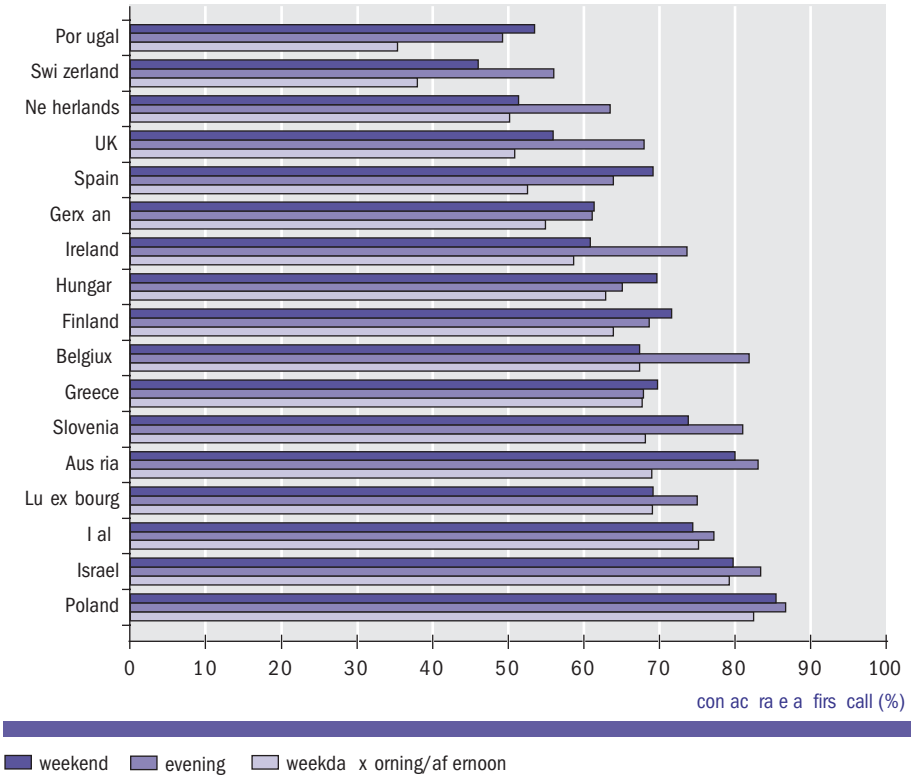
Source: Billet and Philippens (2004), table 1

Billiet and Philippens (2004) took a closer look at the call records and found that interviewers in Germany, Belgium, Ireland, Luxembourg and Austria made on the average less than the specified four calls before giving up a case as a noncontact and ended up with a noncontact rate higher than 3%. They also found that in Ireland, Germany and Belgium, a small percentage of the interviewers were responsible for the insufficient number of calls, whereas in Luxembourg the majority of interviewers broke the "4 calls minimum" rule at least once.

Another indication of differences in the fieldwork hidden by similar outcomes is presented in figure 10.5. This shows the success rates of the first call dependent of the timing of the call. As in section 7.2.1 success is defined here as a contact, and not necessarily as a completed interview. Behind the similar response rates of Portugal and Poland lie two different worlds. In Portugal, the chance of catching someone at home at the first call was below 50%, except on weekends, whereas the success rate of the first call in Poland was above 80%, independent of the timing. The chance

of establishing contact with a sample person is generally smallest during working hours and highest in the evening, although in some countries weekends are even better. In some countries the differences are small (Greece), whereas in the UK the odds of catching someone at home on weekday mornings and afternoons compared with weekday evenings are 75%, 72% in Portugal and 68% in Switzerland.

Figure 10.5 Contact rate at first call for weekday mornings/afternoons, weekday evenings and weekends, ESS round 1 2002/2003



1. Coun ries wi h no or incox ple e call records e cluded
2. Telephone calls in Finland and Swi zerland

Source: Bille and Philippens (2004)

Next to the success rate of calls according to timing, it is important to look at when calls are made to compare the noncontact rate across countries. Billiet and Philippens have shown that interviewers in Ireland, the UK and Spain less frequently make their first call in the evening or during weekends when the contact rate is highest. These countries might achieve a lower noncontact rate or a higher response rate, or minimize the number of calls if they changed their call pattern. Their results show how call records can unearth differences in fieldwork procedures, reveal if and how fieldwork specifications were adhered too and suggest ways for improvements in subsequent rounds.

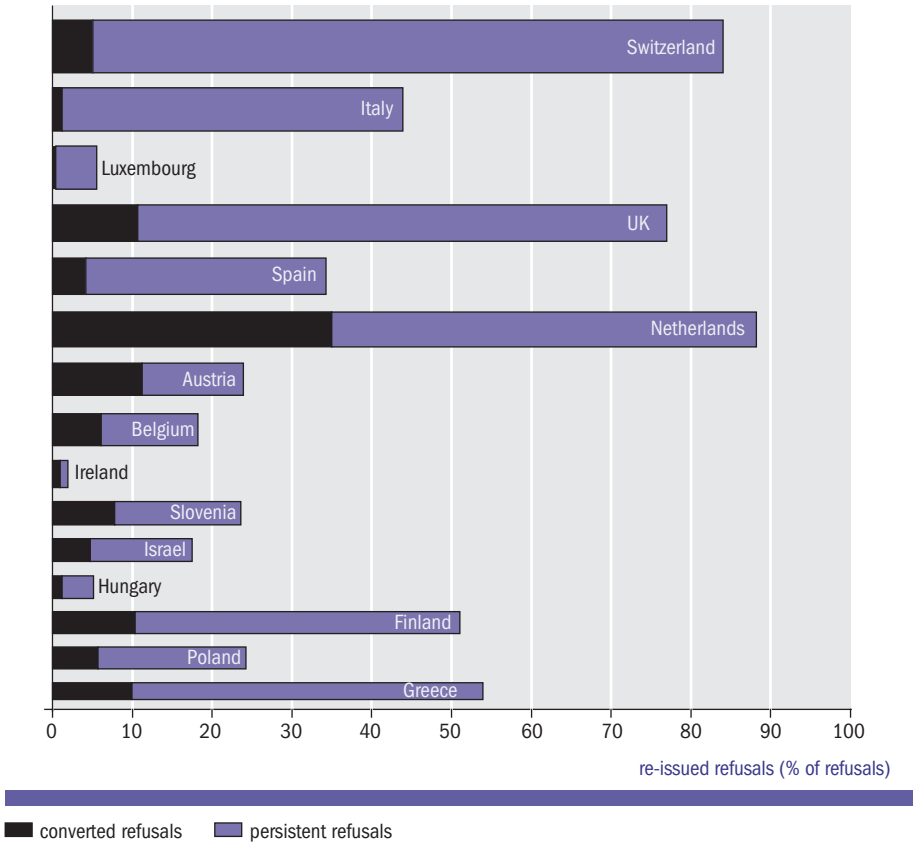
10.5 *Refusal conversion and nonresponse bias in the ESS*

The sections above have shown that there are substantial differences in refusal rates across countries, in type of refusal (by respondent or by proxy), reasons for refusal (Loosveldt et al., 2003) and the deployment of incentives and refusal conversion strategies. Part of these differences may be due to differences in registration. Figure 10.6 focuses again on the data from the call records and shows the results of refusal conversion in countries ordered by final response rates. The width of the bars in this figure represents the final nonresponse rate and the length the percentage of respondents that was re-approached. This figure confirms the sizeable differences in fieldwork procedures between countries. For instance, in Switzerland, the UK, the Netherlands, Finland and Greece more than 50% of the refusals was re-approached with the request to participate, compared to Luxemburg, Hungary and Italy where this hardly ever happened, while response rates were high in Greece and Hungary and low in Luxemburg and Switzerland. Re-approaching refusals substantially increased the response rates in the Netherlands, but had hardly any effect in Switzerland.

Combining the information from figure 10.6 and the previous sections one could conclude that there are high effort, high response countries, high effort low response countries, low effort high response countries and low effort low response countries, at least based on the information from the call records that are complete, which is the case for only a subset of the participating countries. This raises the question of bias due to nonresponse, and especially differences in nonresponse bias across countries due to different field efforts.

How to compare for instance Ireland and the Netherlands, two countries with quite similar response rates (65.4 and 67.9 per cent), taking into account that 23% of the Irish nonrespondents are noncontacts (8.24% of the net sample) compared to 8% of the Dutch nonrespondents (2.55% of the net sample), and that 0.2% of the net sample in Ireland are converted refusals compared with 13.9% in the Netherlands. Sadly, this question cannot be answered. What can be done is to find out whether there are differences between those who cooperated immediately and those who had to be converted. The previous chapters have shown that these are not necessarily similar to the final nonrespondents. Still, one would like to know whether in the ESS the additional field efforts did manage to bring on board respondents who differed from the cooperative respondents. If not, refusal conversion may not be a useful strategy. The study of the effects of refusal conversion is especially interesting in those countries where the majority of refusers were re-approached. In countries where the interviewers re-approached only a small percentage of the initial refusals, one would mainly like to know the procedure that determined whether or not a refusal was contacted again. This information is not available.

Figure 10.6 Refusal conversion, ESS round 1 2002/2003



1. Countries with no or incomplete call records excluded
2. Incomplete information on refusal conversion for Germany and Portugal

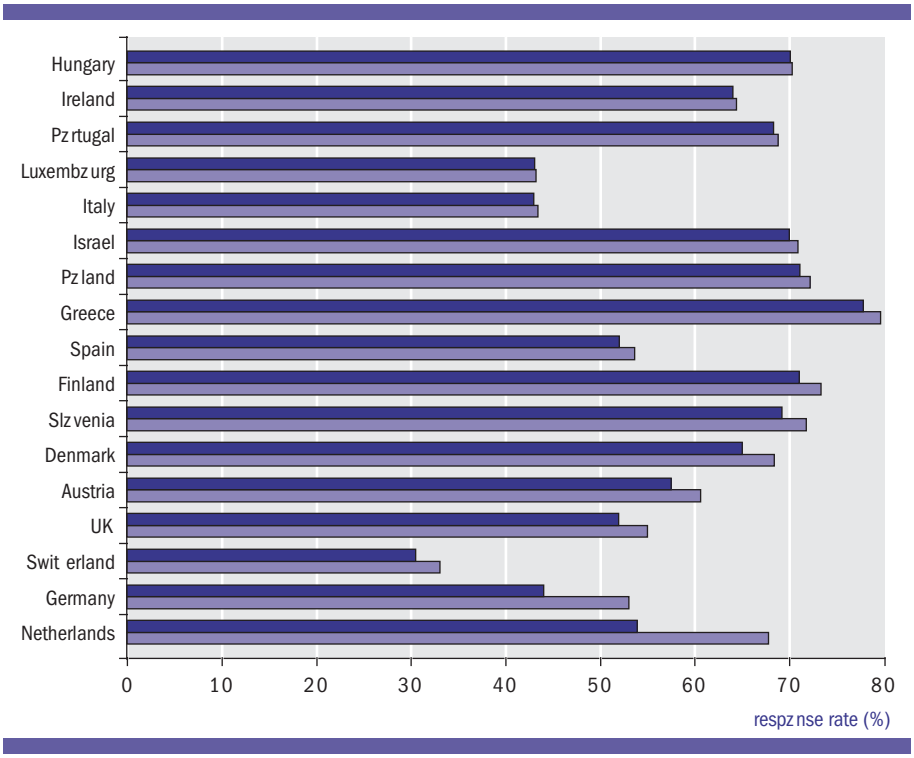
The Netherlands is the most interesting case here for several reasons. Firstly, 88% of the initial refusals were re-approached in the Netherlands. This minimizes the selection effect. Secondly, refusal conversion was fairly successful in the Netherlands, especially considering that not only very promising cases were followed up. As a consequence, refusal conversion significantly increased the response rate and yielded a fair number of converted refusals. In addition, the ESS fieldwork in the Netherlands was carried out by the same organization that conducted the fieldwork of the AVO, analysed in the previous chapters. In trying to obtain cooperation from initial refusal on the ESS they used the experiences with the AVO follow-up survey among persistent refusers, as described in chapter 6 and 9. And finally, the results of refusal conversion in the Netherlands in two different surveys can now be compared.

Billiet and Philippens (2004) have studied the results of refusal conversion for those countries that yielded a fair number of converted refusals, namely Austria, the UK, Switzerland, Germany and the Netherlands (see also figure 10.7 where countries are ranked according to the relative yields of refusal conversion). One interesting result of their study is that the results for Austria are all contrary to expectations: converted respondents in Austria do not live in big cities and they have a higher rate of political participation, even after taking background variables into account. One explanation might be that only a small part of the refusals were re-approached in Austria (see figure 10.6) and that this subgroup do not represent the refusals. For Switzerland and the UK the differences between cooperative respondents and converted respondents were generally not significant, most probably due to the small sample size. Most information can be obtained from the converted refusals in Germany and the Netherlands. These were subdivided into 'soft' and 'hard' refusals. Soft refusals were defined as initially reluctant respondents who cooperated at the first contact following the initial refusal. Hard refusals needed several subsequent contacts and additional incentives (in the Netherlands). Billiet and Philippens found that political trust and political participation were lower among the initially reluctant respondents, and the same applied for social trust in Germany. These differences were attenuated after incorporating background variables, and were larger for hard refusals than for soft refusals.

Another study of the Dutch ESS nonresponse data[5] yielded similar outcomes. In this case the distinction between soft and hard refusals was based on interviewer assessments, in a similar way to the approach in section 8.3.1. Those refusers who, according to the interviewer, would definitely not cooperate in the future were labelled "hard" refusals, while those who would probably not cooperate, might perhaps cooperate or would certainly cooperate were labelled "soft" refusals. An analysis based on this distinction showed that "hard" refusals were on average less politically and socially integrated, had a less favourable attitude towards immigrants and Europe and its institutions, were older and less likely to have high incomes compared to cooperative respondents. 'Soft' refusers, on the other hand, were rather similar to cooperative respondents.

5 Michel Philippens and Ineke Stoop (2004) *The effects of selective refusal conversion strategies on the reduction of nonresponse bias*. Unpublished manuscript

Figure 10.7 Response rates before and after refusal conversion, ESS round 1 2002/2003



1. Czuntries with nz zr inczmplete call reczrds excluded
2. Data fzr Swit erland based zn subsample with telephzne recruitment

Szurce: Billet and Philippens (2004)

These results indicate that in the Dutch ESS, contrary to the AVO outcomes in the previous chapters, temporary refusals differ from immediate respondents and are similar to the stereotypical image of refusals as showing less (political) participation. Whether this is due to the fact that more refusals were re-approached, or to the fact in the ESS that the topic of the survey might be more related to cooperation than in the AVO, or to other reasons, is not clear. It is also not clear whether the temporary refusals in the Dutch ESS are similar to the final refusers, as there was no follow-up survey among persistent refusals to give more conclusive information on possible bias. And finally, it is definitely not shown that the dissenting opinions of Dutch temporary refusals are shared by temporary refusals in other countries.

10.6 *Contact forms: treasure trove or can of worms*

This chapter has been included to reflect on the results of the previous chapters. The main message of this chapter was to show how close monitoring and detailed recording of the fieldwork process can highlight differences between countries that would otherwise have remained hidden, show how fieldwork can be improved and point to possible bias in survey outcomes. The contact forms of the AVO and the ESS can be considered an important instrument in achieving these goals. This chapter and chapter 6 have also shown that the use of contact forms is not without its problems. Firstly, keeping and keying call records takes time, and keeping them according to the guidelines is not something that happens a matter of course, especially within organizations that are not used to doing this. Experiences with the AVO1999 and ESS2002/ 2003 contact forms have in both cases led to improved forms in a future round; more complete for the AVO and simpler for the ESS. Furthermore, data protection regulations and privacy laws may obstruct data collection on nonrespondents and the response process. In addition, keeping call records may be more attractive, deliver better quality data and possibly lead to more efficient fieldwork procedures, if this is not just an extra for other people to analyse, but a tool that can help monitoring fieldwork and adjust procedures during fieldwork. Finally, the results in this chapter and in the previous chapters show that even with standardized contact forms and detailed call records, no information is available on the decision whether or not to re-approach a (temporary) refusal again.

As was stated in section 2.6.3, simply re-approaching promising cases to enhance response rates may not be the best strategy. The results for Austria illustrate this. The assumption that 'difficult' respondents are similar to final nonrespondents was rejected in the previous chapters. As in chapter 8, this chapter has tried to distinguish between (soft and hard) converted refusals, presumably because they put up different amounts of resistance. The predictive values of interviewer assessments of the likelihood that a refusal would cooperate at a future contact were discussed in section 2.6.3. Indeed, Infas (2003), the organization that conducted the ESS fieldwork in Germany, observed that interviewers may not be the best judges of future respondent cooperation. An even more difficult problem concerns the first distinction between soft and hard refusals in this chapter, namely the number of contacts and possibly additional incentives. As was mentioned before it is quite likely that these extra efforts will be deployed only if future success seems likely, which brings us back to conclusion formulated in chapter 8 that it will not be possible to equate the amount of effort needed to persuade a respondent to cooperate with the amount of resistance put up by the respondent.

Figure 10.8 Types of respondents and refusals

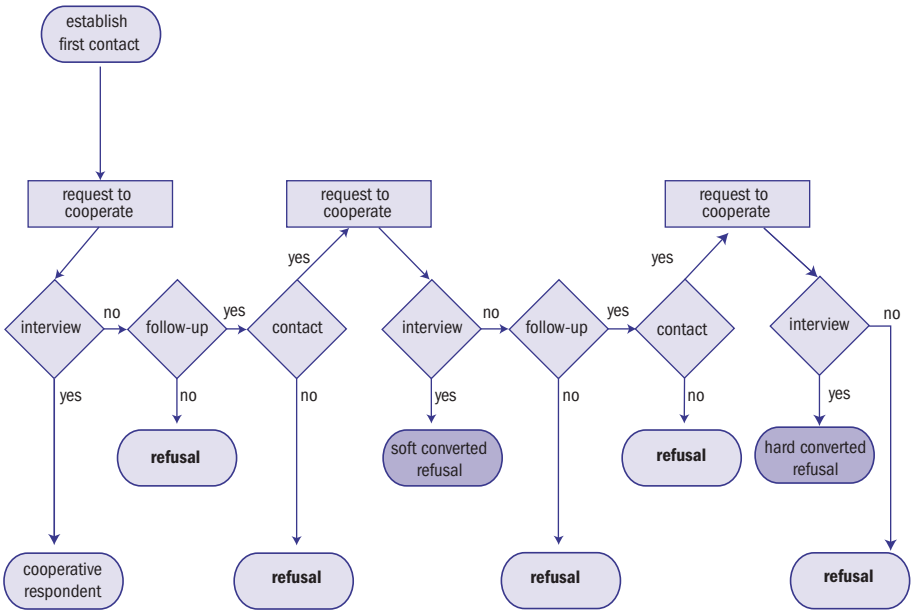


Figure 10.8 contains a slightly amended version of the model presented in figure 2.5 and represents what happens after first establishing contact with a potential respondent. It shows that for a respondent to be classified as a soft refusal who finally cooperates, 1) the respondent should initially refuse, 2) the fieldwork organization should decide that this refusal should be re-approached, 3) the refusal should be contacted again and 4) the refusal should be converted. For a 'hard' converted refusal there are even more steps that can partly be chalked up to the initial reluctance of the respondent, partly to organizational decisions and partly to the success of efforts to re-contact an initial refusal again. These decisions and fieldwork steps are not independent. Figure 10.8 also shows that there are many types of final refusals, namely those that are not re-approached after a first refusal, those that cannot be contacted for the second time, those that are not re-approached after the second refusal, those that cannot be contacted for the third time, etc. etc. In a survey where no attempts at refusal conversions are made, these different types of refusals of course cannot be separated.

As in the previous chapters, this overview of nonresponse in the ESS supports the need for contact forms and call records, despite their shortcomings, in order to

unravel the nonresponse process, improve fieldwork, compare countries and estimate bias. The outcomes for the ESS differ, however, from those of the AVO. In the latter case there were very few differences between cooperative respondents and temporary refusals, whereas the results of the ESS indicate that converted refusals may participate less politically. This may be due to the subject of the survey, as the AVO is focused on social participation rather than political involvement, or to the more intensive efforts at refusal conversion in the ESS. It is possible that the 'hard' refusals as identified by Billiet and Philippens (2004) are similar to the persistent refusers who cooperated in the follow-up survey of the AVO. Whatever the reason, it seems useful to repeat the concluding recommendation from Billiet and Philippens that some topical information should be obtained from the nonrespondents in order to test for nonresponse bias. They suggest either a central question procedure (but what is the central question in the ESS?) or a follow-up telephone survey of a subsample of persistent refusals. Both approaches will be expensive and difficult to implement in a cross-cultural survey. On the other hand, one might argue that in a survey aimed at cross-cultural comparison, differences in fieldwork approaches and types of bias across countries might be a serious risk.

11 *Enhancing response rates: how and why*

The preceding chapters have shown how high response rates can be achieved and how the effect of increased response rates on the accuracy of survey estimates can be measured. Based on the literature and on the empirical evidence, in this chapter the six central questions formulated in section 1.3 will be revisited. Preliminary answers were given in section 5.5, based on an inventory of the research literature.

Q1 Why are high nonresponse rates a reason for concern?
Q2 Who are less likely to respond, either because they are more difficult to contact or because they are more reluctant to cooperate?
Q3 How can response rates be enhanced?
Q4 How to study nonresponse?
Q5 Do enhanced response rates improve the accuracy of survey outcomes?
Q6 How to combat nonresponse error and allocate funds effectively?

These questions can be divided into two parts. Firstly, an inventory of nonresponse studies could give a semi-final answer to the first three questions. In this case, general findings and good practices from the literature have been used to implement fieldwork and to develop an analysis scheme for the empirical study in the last chapters, based on the Dutch Amenities and Services Utilisation Survey (AVO) 1999 introduced in chapter 6 and the European Social Survey (ESS) 2002/2003 introduced in chapter 10. In the next section it will also be shown how the first three questions can be answered with regard to these two particular surveys. Secondly, after the study of earlier nonresponse research the three last questions remained on which the literature there is less than complete agreement among nonresponse researchers or no conclusive answers and complete guidance could be provided. In order to answer the final question 'How to combat nonresponse error and allocate funds effectively?' instruments were developed to measure nonresponse (Q4) in order to test if enhanced response rates do reduce nonresponse error (Q5).

This chapter addresses the central questions of this study based on both the nonresponse literature and the empirical studies presented in chapters 6 to 9. Earlier research has already been summarized in section 5.5, which is why this chapter will focus on new evidence. In this chapter the second question will be answered before the first question, because the present evidence on response selectivity (Q2) will be taken into account to decide whether nonresponse is a reason for concern (Q1) in the AVO and the ESS.

11.1 *Who are less likely to be contacted and to cooperate?*

The results of the AVO study on contactability presented in chapter 7 are in line with the literature summarized in section 5.5.1. Persons living in urban areas, who have activities outside the home and who are part of smaller households are hard to contact, as are households without a listed telephone number, which is common among urban dwellers, ethnic minorities and younger persons. Activities outside the home are indicated by general background characteristics such as being in paid work or a student, but can also be leisure activities such as attending cultural events. The effects should be considered as additive: urbanicity, being single, not having a listed phone only and attending cultural events will be correlated but all these effects also add up to make a person increasingly hard to reach.

The AVO results for reluctance to cooperate are more difficult to interpret. According to chapter 8, respondents living in big cities were less likely, while men, singles, members of ethnic minorities and the higher social strata were more likely to cooperate immediately. However, it cannot be concluded that big-city dwellers were uncooperative and men, singles, etc. more cooperative, as this relates to respondents only. Among big-city dwellers there were many temporary refusals because they cooperated less at the first contact and because a large proportion of them did cooperate in the end. On the other hand, among single males there were few temporary refusals because few cooperated at the first contact and only a small proportion of initial refusals were converted. The final refusals differed from the respondents in ways that could be expected, considering the results of refusal conversion. They were slightly more often male and single and less often couples with children. Due to the extended field efforts in big cities, the final refusals actually less often live in urban areas.

Interestingly, cultural participation was also related to reluctance to cooperate. Whereas those who attended cultural events of whatever type were harder to reach, converted refusals comprised more persons who preferred classical culture only (theatre, concerts), whereas final nonrespondents more often attended popular cultural activities only, such as going to the cinema, pop concerts and dance events. Final refusals also performed less sport, less often had a religious affiliation and less often owned a PC. If they did, they were more active on the Internet, however, and used the PC more for chatting and games rather than for serious purposes. Very similar results were found by Pääkkönen (1999) and Väisänen (2002) in their analyses of the Finnish time budget survey (section 3.3.5). There are two possible explanations for these effects. Firstly, one could see converted refusals as responsible, cultural mainstream citizens who may just need some prodding to cooperate, and final refusals as being less part of the cultural mainstream. This would be in line with Groves and Couper's (1998) theory on social isolation. They assume that people who are socially involved in their society have more to give to and more to receive from

this society. People who are not in touch with the mainstream culture of a society might be more prone to reject survey requests. The effects are not very large, however, and stigmatising filmgoers as not being in touch with the mainstream culture may be taking things a little too far. The second explanation is that a survey of the utilisation of amenities and services may have less appeal for persons with a youthful lifestyle who may not regard themselves as users of those amenities and services, despite the fact that these include pop concerts, dance events and art house films.

In summary, the AVO results on contactability conform to the literature. The AVO results on cooperation show substantial differences between temporary refusals and final refusers. It should be added, however, that only a small number of the variables in the AVO showed an affect for contactability and reluctance to cooperate and that the results were fairly weak, except for PC ownership and internet use.

The European Social Survey nonresponse data allow a comparison to be made between cooperative respondents and reluctant respondents, but no information was available on final refusals. The results presented in chapter 10 show that in the Netherlands political trust and political participation are lower among the initially reluctant respondents, and that temporary refusals from whom the interviewers did not expect future cooperation were on average less politically and socially integrated, had a less favourable attitude towards immigrants and Europe and its institutions, were older and less likely to have high incomes compared to cooperative respondents. Less adamant refusers, on the other hand, were rather similar to cooperative respondents. As these results are indicative of nonresponse bias, they will recur in the next section.

11.2 *Is nonresponse in the AVO and the ESS a reason for concern?*

Nonresponse in the AVO turned out to have only a minor effect on the accuracy of estimates (see section 9.4). Differences in environmental and socio-structural characteristics and survey variables between respondents and nonrespondents were small. After taking these small compositional effects into account, it turned out that substantive differences could only be found for a very small number of survey variables (see previous section). These results may of course be specific to this round of this survey.

There is more reason for concern with regard to the ESS (chapter 10). What is especially worrying is that the response rates differed immensely between countries, from 33% in Switzerland to 80% in Greece. Whereas in theory, nonresponse bias in Switzerland could be lower than in Greece, the maximum size nonresponse bias can assume is of course much higher in Switzerland (see section 2.3). A second reason for concern is that the composition of the response differs between countries. Section 10.4 shows that the noncontact rates among countries varies substantially

and that there are vast differences in refusal conversion approaches, ranging from re-approaching virtually no (initial) refusals (Ireland) to re-approaching almost all refusals (the Netherlands). From the body of literature presented in chapters 3 to 5 and from the analyses of the AVO it can be assumed that late contacts differ from early contacts. Differences in contact rates between countries can therefore have an impact on cross-cultural differences. The effect of converting refusals was traced for a number of countries in which this procedure yielded a substantial amount of additional respondents. The results indicate that the initially most reluctant respondents in the Netherlands and Germany scored lower on political trust and political participation and were on average less politically and socially integrated (see previous section)

These results differ from the results of the AVO, which did not show any substantive differences between cooperative respondents and initially reluctant respondents. This might be due to the difference in topic, i.e. factual questions in the AVO versus questions on opinions in the ESS. From the AVO it was concluded that temporary refusals cannot be used as proxies for final nonrespondents. The differences between easy respondents and converted refusals in the ESS cast doubt on whether this conclusion is generally applicable. No independent information on final refusals was available in the ESS, which makes it impossible to give a conclusive answer. Billiet and Philippens (2004) interpreted their ESS results as indicative of nonresponse bias in the Netherlands and Germany. Other countries were bias-free, simply because there was no information on final refusals and no information to distinguish reluctant from cooperative respondents and thus to estimate bias.

This leaves us with the final conclusion that nonresponse in the AVO is not very worrying, but nonresponse in the ESS should be a serious concern because of differences in response rates, differences in response composition and differences between cooperative and initially reluctant respondents in those countries where these could be measured.

11.3 *How were high response rates achieved: unravelling the process*

General strategies to enhance response rates were presented in section 4.3, and particular strategies to obtain a high response among a very difficult group, namely adamant refusers, in sections 6.3 and 9.2. This section will look strictly from the outside and attempt to give an indication of the contribution of different steps in obtaining response, namely minimizing the noncontact rate, making appointments with respondents, resuming broken-off interviews or completing incomplete sets,

re-approaching households that were recorded as not able to be interviewed ('situational refusals') and converting downright (temporary) refusals.[1] This section is based on the response process of the AVO, but where relevant results from the ESS will be added.

The baseline: minimum efforts

What would the response rate of the AVO99 have been in a minimum effort variant? If only interviews conducted at the first call had been taken into account, the response rate would have been 15.7%. A more realistic estimate of a minimum variant is based on a maximum of four calls to previously noncontacted addresses and interviewing those households only that are willing to be interviewed on the spot. Assuming an 8.5% noncontact rate after four calls (based on the timing of first contact in chapter 7) and according to the outcomes of first contact in table 8.1 and 8.2, the response rate in this minimum variant would have been 32%.

Maximizing the contact rate

In the AVO the maximum number of calls to previously noncontacted addresses was nine rather than four, which is not an unusual number (see section 6.2.2). This would increase the contact rate from 91.5% to 98.4% and the response rate (based on first contact interviews only from the baseline) from 32% to 35%. Chapter 7 described how this high contact rate has been achieved. Besides making many calls, interviewers called at different times of the day and on different days of the week. The discrete hazard rate logistic regression model of the probability of contact, adapted from Campanelli, Sturgis and Purdon (1997) confirmed the importance of making evening calls.

In their analysis of the ESS Billiet and Philippens (2004) showed that – at least in Ireland, the UK and Spain – the number of calls could have been smaller and the contact rate would have been higher if more evening or weekend calls had been made. Two reservations should be made with respect to call efficiency. Firstly, evening and weekend calls may have a higher contact rate, but may also be more expensive and the number of evening and weekend hours during which calls can be made is limited.

1 Isolating these steps is somewhat artificial, for instance because the timing of calls might have differed if the maximum number of calls had been set a low level right from the beginning (see section 5.2). In addition, the disposal of multiple households is difficult to follow across these different steps (see annex 6.1). Furthermore, if there would have been no occasion for following up on appointments, interviewers might have been more keen on conducting the interview on the spot, and finally, the contribution of different steps is based on the outcome of the first contact whereas subsequent contacts can have different outcomes (see table 8.3). For this reason, the contribution of the different steps should be seen as an approximation.

Secondly, interviewers may have their own call efficiency optimization strategy that is not visible from their call records. They may call on a respondent when they are in the neighbourhood anyway, and thus may make a very inexpensive call, or base their timing of calls on local knowledge or knowledge from a previous unsuccessful call.

After four unsuccessful calls in the AVO, interviews were reassigned in a second wave of the fieldwork and after that in a third wave (see section 6.2). At the end of each wave interviewers could leave a card or make a telephone call to the sample household if a telephone number was available.[2] It is often assumed that telephone contacts result in a higher refusal rate (see section 4.3.3). Regarding hard-to-reach respondents, however, the choice is between a high probability of noncontact and a possibly higher refusal rate. In the AVO the latter risk was minimized because (telephone) refusals were often re-approached in order to persuade them to participate (see below). This combined approach has managed to achieve a very low noncontact rate, which is of course the basis of a high response rate.

The analyses of contactability in the AVO1999 led to improved contact forms and call records (see section 11.4) in the AVO2003. As a consequence, the contact rate, and thus the response rate in 2003 decreased substantially (GfK, 2003). This was mainly due to the fact that in the AVO1999 addresses on which no one had been found at home after nine calls were recorded as vacant and thus ineligible (see section 6.2.2). One of the recommendations of the present study, the collection of independent information on the vacancy of units (see section 6.2.3), was implemented in the AVO2003, which resulted in a noncontact rate of 4.3%, compared to an estimated 1.5% in the AVO1999. This shows that it is not necessarily the case that using information from call records improves response rates in future surveys, or at least not published response rates. It does show why it is strongly recommended that paradata be collected and call records kept: it at least produces more transparent response results.

Maximizing the cooperation rate: survey design and response target

General steps in order to achieve a high response rate in the AVO comprised careful preparation of the survey (advance letter, brochure), and the hiring and training of experienced and motivated interviewers. In the Netherlands, interviewers are sometimes instructed to achieve a target number of interviews from their assignment. This is similar to the first fieldwork target described by Groves and Couper (1998, p. 38-39), similar to quota sampling, which is maximizing the number of achieved interviews per time unit, if only because interviewers have a limited time for con-

2 Section 7.2.3 showed that the availability of a listed telephone number has a strong middle class bias in the Netherlands. This is unlikely to have influenced the AVO outcomes, however, as households without a landline telephone or ex-directory received a house call anyway.

ducting interviews: 'An acceptance of the survey request is preferred to a denial, but a lengthy, multicontact preliminary to an acceptance can be as damaging to productivity as a denial.' In the AVO and the ESS they were told to try to obtain cooperation from all households they had been assigned, which is similar to the second fieldwork target of Groves and Couper which is minimizing refusals: 'The odds of success are increased with the continuation of the conversation. Thus the interviewer does not maximize the likelihood of obtaining a 'yes' answer in any given contact, but minimizes the likelihood of a 'no' answer over repeated turntaking in the contact.'

In the European Social Survey (see chapter 10) a target response rate of 70% was set. Suggestions on how to achieve this were presented in a document drafted by Achim Koch from ZUMA, one of the members of the Central Coordinating Team (ESS, 2004). Rather than 'just' aiming for high response rates, participating countries were warned that they should be '... mindful of the need to boost levels of response amongst all groups of the population and to bring response rates to a more consistent level across subgroups, if possible'. Survey cooperation should not be limited to specific groups: 'Certain elements of the survey design may differentially affect the likelihood of participation among different groups of the population. For instance, a monetary incentive may be more likely to encourage the participation of people with low incomes rather than those with high incomes. Measures to reduce nonresponse should take account of such issues, targeting groups who are disproportionately underrepresented as a result of design issues.'

In survey fieldwork, the first blow is half the battle and cooperation obtained during the first contact certainly makes life easier. A 'light' refusal conversion may have occurred during the first contact when the interviewer persuaded a reluctant sample person to cooperate anyway. If immediate interviewing is not possible, the steps mentioned below can be taken, namely making appointments, resuming broken-off interviews and converting refusals. What happened during the first contact would only be visible if information on the doorstep interaction had been recorded (see section 4.2.2). The fact that a refusal has been converted during a subsequent contact will be visible from the contact form (see section 11.4) because it requires additional calls.

Maximizing the cooperation rate: making and keeping appointments

In order to simplify assumptions, from now on the cooperation rate rather than the response rate will be taken into account (see figure 2.2). To increase the cooperation rate, appointments can be made and should be meticulously followed up. This requires experienced interviewers who are motivated, can convey the importance of the survey, know when to push and when to withdraw, and who will conscientiously keep appointments. These combined actions would increase the cooperation rate (based on a 98.4% contact rate) from 36% to 49%.

Table 8.1 showed that 14.3% of first contacts in the AVO resulted in an appointment and table 8.2 showed that 82% of these appointments resulted in a successful interview. All appointments were followed up. This should be a matter of course, but experiences with other surveys have shown that this does not always happen. Not keeping appointments may be due to interviewers with too high a workload or too short a fieldwork period. One of the reasons why appointments should always be kept is that not doing so shows disrespect for respondents.

Having an interview preceded by an appointment may be a successful strategy, firstly because of interviewer logistics. Interviewers may make many calls in the early evening, when respondents are most likely to be at home, and then set up an appointment for another time. Of course, if the respondent is willing to be interviewed at the first contact, it may be risky to defer this to a later time. The second reason why the high rate of appointments may reflect a successful field strategy is that experienced interviewers are well-versed in tailoring and maintaining interaction (section 4.3.5). If the respondent does not appear to be to keen on being interviewed, a good interviewer might withdraw and suggest making an appointment at a more opportune moment. This may make it more difficult for the respondent to refuse outright (although 18% of the appointments made at the first contact did not result in an interview).

Maximizing the cooperation rate: completing broken-off interviews and incomplete sets

According to tables 8.1 and 8.2, 5.3% of the first contacts with eligible cases in the AVO resulted in a broken-off interview. One third of these households finally delivered a complete set which means a net result of 1.7%. The AVO suffered from a rather high proportion of incomplete sets, due to the fact that a household was considered as a respondent only if all household members aged 6 years and older had completed a drop-off questionnaire and if the face-to-face questionnaire with a responsible adult had been completed. According to table 8.2, 5.7% of the contacted addresses ended up with incomplete documents, generally comprising a completed interview but no filled-in drop-off questionnaires.[3]

3 In substantive analyses the incomplete sets are ignored, as the AVO is a household survey and as the face-to-face interview comprises only a small share of the information to be collected (see section 6.2.1). The information from these incomplete sets could have been used more effectively (to measure or adjust for nonresponse bias) if the face-to-face interview had comprised slightly more information, also pertaining to other household members, comparable to the questionnaire of the follow-up survey. This is not simple, as this would lengthen the questionnaire by including questions that would fit better in the drop-off questionnaire anyway. An alternative might be to ask a small number of additional questions to the partially compliant respondent after it is clear, at the end of the interview or when the interviewer comes back to collect the questionnaires, that no drop-off questionnaires will be filled in.

With hindsight, it is a pity to dismiss this partial nonresponse because it does not contain the required information on the entire household. For substantive research purposes, information on the complete household had to be available, but for nonresponse measurement cooperation from one household member only could be of great use, especially if this person gave some additional information on other household members. This would mean introducing a different questionnaire halfway through, when complete cooperation appeared not to be feasible, which might create logistical problems. An additional concern is that interviewers might introduce this short household version too early and that complete response would actually decrease. A practical problem is that it is generally only clear after completion of the face-to-face interview, and possibly only when the interviewer comes by to collect the completed questionnaires, that no or not all drop-off questionnaires are filled in.

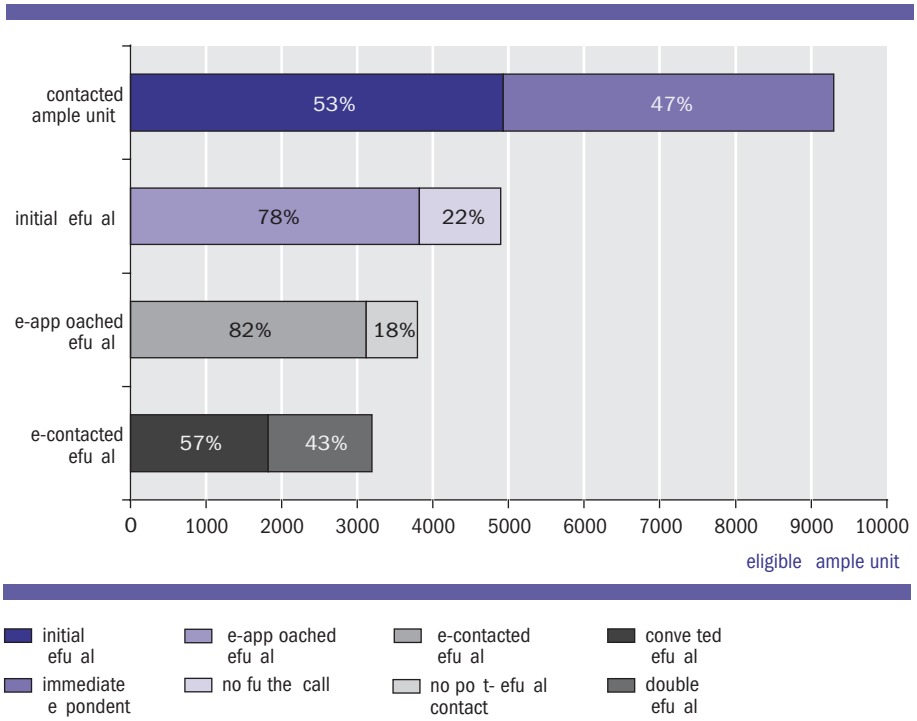
Maximizing the cooperation rate: converting refusals

Converting refusals raised the cooperation rate in the AVO from 49% after a follow-up of appointments and completing incomplete sets to 66%. In the AVO about one third of the refusals at the first contact finally cooperated, as did one third who said they were not able to cooperate. Section 2.6.3 has shown that revisiting initial refusals is not a standard procedure in many countries and is sometimes even not allowed. Results from the ESS presented in section 10.5 confirmed this and showed the wide variety in the number of refusals that were re-contacted and in re-contacted refusals that were converted. Converting refusals can be subdivided into two parts, namely re-approaching those who said they were not able to participate when the interviewer first came by and converting initial refusals.

Figure 11.1 presents a summary overview in which refusals and 'not ables' are jointly called refusals.[4] It comprises the outcomes of all 9261 contacted households in the AVO. Of these, 47% complied immediately, which includes completing in interview after an appointment or breaking it off. The other 53% at least once showed non-compliance (refusal or not able). Of the 53% initial refusals, 22% received no further calls (including the incomplete sets). Of the 78% re-approached initial refusals, 18% were not contacted again, which means that they were not at home when the interviewer called after a refusal.

4 Previously in this section final outcomes and the outcome of the first contact were related, whereas figure 11.1 traces what happens with households who ever refused, irrespective of the contact at which this refusal occurred. The outcome of the first contact of a temporary refusal in figure 11.1 could be an appointment, followed by a 'not able', followed by a second appointment and finally followed by an interview.

Figure 11.1 Dealing with refusals in the AVO 1999



And finally, of the initial refusals who were re-approached and were found at home, 57% finally agreed to cooperate at the second or a later contact. As mentioned in chapter 8, those refusals who were not re-approached or who were but were not found at home, may differ from those who received a second request for participation. The 57% cooperation rate among re-approached, re-contacted refusals is still remarkably high and seems to indicate that
many reluctant respondents do not see a second request as a gross infringement of their privacy but more likely as an indicator of involvement of the interviewer, the fieldwork organization and the sponsor.
In addition to the fact that a quite substantial number of temporary refusals was converted, the question remains as to which types of respondents and reasons for refusal are susceptible to refusal conversion in a regular survey. The results of the AVO are fairly disappointing. From section 11.1 it can be concluded that the success of refusal conversion was greatest among those groups who were cooperative anyway which resulted in a more unbalanced composition of respondents than before refusal conversion, and no change in substantive variables. The results of the ESS were somewhat more promising as, at least in two countries, refusal conversion seems to have brought on board extra respondents who met the profile of probable (final) refusals.

In summary, it seems that Lievesley's (1983) strategy of re-approaching only those refusals who according to the interviewer are promising of future success (see section 2.6.3) should not be recommended. This might be a cost-effective way of enhancing response rates but, as can be judged from the evidence of the AVO and the ESS, interviewers may not be the best judges of respondent cooperation, and re-approaching soft refusals only might do little to reduce nonresponse bias. Although it may neither be possible nor advisable to re-approach outspoken, hard refusals as a general procedure, at least the selection process of who will be re-approached should be made transparent in order to measure nonresponse bias.

Maximizing the cooperation rate: lessons from the follow-up survey among refusals

Re-approaching outspoken, hard refusals – although with a substantially smaller response burden – was exactly what was done in the follow-up survey among persistent refusers in the AVO. How to obtain cooperation from this group was described in some detail in section 6.3. The best interviewers were selected, who were made very well aware that they were part of an important scientific endeavour, well paid, well briefed and closely supervised and guided, given all means imaginable and who had the total commitment of sponsor, survey organization and survey staff. No expenses were spared. Indeed the limited information from each respondent of the follow-up survey (15 minute face-to-face interview) was about twice as expensive as the extensive information obtained from each completed household in the regular survey (half hour questionnaire, very long self-completion questionnaire for every household member of 6 years or older). With interviewers so highly motivated and so well supported very high cooperation rates are possible. The response-enhancing strategies from the follow-up survey in the AVO were implemented in a somewhat less extensive way in the refusal conversion phase of the Dutch ESS. One concrete measure that was adopted was giving an interviewer a certain amount of cash (EUR 22.5 five years ago in the AVO follow-up; EUR 15 in the Dutch ESS 2004/2005 in the final phase of the fieldwork for the most reluctant respondents), to be spent as they deemed fit.
If the procedures with regard to incomplete sets had been focused at collecting crucial data in a 20-minute interviews from partially reluctant households, if all persistent refusals had been re-approached, instead of a small sample only in the follow-up survey, the final response rate could have been 90% (66% response + 5.7% incompletes + 72% of 25% final refusals). The results of Voogt (2004), presented in section 5.3, confirm that very high response rates can be achieved provided one is willing to accept a short questionnaire and to adapt to the preferences or even pander to the whims of the sample persons.

Maximizing the cooperation rate: irritation and satisficing?

A frequently expressed concern is that putting (polite) pressure on respondents to cooperate may harm data quality The results of the AVO presented in section 8.4 gave no indication that refusal conversion or re-approaching persistent refusals leads to a lower data quality as measured by item nonresponse. Section 9.3.1 gives

some indication that telling interviewers that some sensitive questions (in this case on income) do not have to be answered may increase unit nonresponse. It should be considered whether this trade-off between unit nonresponse and item nonresponse is worthwhile. In this case income was an essential survey variable.

11.4 *How to study nonresponse?*

In section 5.5 the two models were introduced to assess the impact of nonparticipation on survey estimates in the subsequent chapters, namely the continuum of resistance model and the classes of nonparticipants model. The results of the tests of these models will be presented in the next section. In this section the focus is on the instruments with which to collect the data to test these models. Two instruments were developed and implemented to measure whether and how nonresponse causes bias and enhanced response rates may reduce bias, namely the contact forms and a follow-up survey among persistent refusals in the AVO. This section summarizes the design, purposes and problems of these contact forms and then discusses the intricacies of measuring contactability and reluctance to cooperate. The design of the follow-up survey among adamant refusers to the AVO was described in section 6.3, which showed how a high response rate in this survey was achieved. This section also showed the problematical aspects of a follow-up survey, including the period elapsed between the regular survey and the follow-up, the adaptations in the questionnaire, the sampling procedure, the type of sample (one person instead of the entire household) and the high costs. The control group was required to meet a number of these problems. The follow-up survey did provide information with which nonresponse bias could be measured and with which temporary refusers could be compared with final refusals. This, and the high response rate mean that it can be considered a useful tool for measuring bias.

From the literature presented in section 2.6.4, section 6.2 on the design of the AVO and chapter 10 on the ESS, it will be clear that contactability and reluctance to cooperate can only be measured by detailed information on the field history of individual sample units. The form on which this information can be recorded during fieldwork is called a contact form, although it also comprises complete call records and thus also information on calls where no contact was made. The contact forms used in the AVO were presented in Annex 6.4 and 6.5 and discussed in section 6.2. The ESS contact forms (sections 10.3 and 10.6) for different types of sampling frames are available on the ESS website (www.europeansocialsurvey.org, see ESS docs). These forms make it possible to assess the contactability and inclination to cooperate of each sampling unit, provide reasons for refusal, give assessments of possible future success, etc. This information can be used to estimate the impact of nonresponse and the possible size of nonresponse bias and to correct for nonresponse. In addition, as chapter 10 has shown, information from these contact forms can be used to assess whether fieldwork has been carried out according to the guidelines. Ideally, the information should also be used

during fieldwork, to provide information on timing of individual contact attempts based on earlier attempts and to regulate the refusal conversion strategy. To do this, the information from the contact form should be on-line available, instead of being processed after the end of fieldwork (see Groves and Ziniel, 2003). If this were so, the contact forms could also play a role in response enhancement and therefore maybe be more attractive to field organizations and interviewers.

Contact forms can only capture a rough approximation of what really went on in the field. The analysis of contactability in section 7.2.1 showed that even with detailed call records contactability is difficult to measure because interviewers may optimize the probability of contact with a hard-to-reach respondent by basing their call patterns on local knowledge or information from earlier calls. Someone may be very hard to reach and be rarely at home, but when the neighbour says that she is likely to be at home on Thursday morning when the chimneys are being swept, she will suddenly be easy to contact. On the other hand, interviewers may make calls at fairly unpromising times on the off-chance that someone is at home when they are in the neighbourhood anyway. This would decrease contactability (another unsuccessful call), but would add very little to fieldwork costs.

Knowledge on how reluctant sample persons are and why they are reluctant is of use in obtaining cooperation, deciding who to re-approach after a first refusal, measuring nonresponse bias and adjusting for nonresponse. Measuring reluctance to cooperate is much more complex than measuring contactability (section 8.1). In this study three indicators of reluctance have been included, namely reasons for refusal, interviewer assessments of the likelihood of future success after a refusal and finally the number of (temporary) refusals. The time and effort it took to persuade the prospective respondent to cooperate can also be measured by recording doorstep interactions on tape or paper forms (see section 4.2.2); this was outside the scope of this study.
Not every respondent may give a reason for refusal, and interviewers seem to be most vexed by persons who just slam the door in their face and by people who politely refuse without giving an explanation. Reasons for refusal can be subdivided into situational and survey-related reasons (section 4.2.2), of which the former are more susceptible to refusal conversion. Section 8.2.1 showed that reasons for refusal may only partially reflect why persons do not comply, because unwilling sample households gave different reasons for refusals at subsequent contacts. For instance, of those who said they were too busy to participate in the AVO at the first contact more than a third plainly refused at a subsequent contact. This gives the impression that people may just try to find an easy way out, or mention the first thing that comes to mind in order to get rid of the interviewer.

Interviewer assessments of the likelihood of future cooperation of initial refusals are an imperfect indicator of reluctance (see section 5.2). Table 8.4 showed that those who were characterized as absolutely non-cooperative in the AVO were re-

approached in 70.5% of cases, of whom almost half ultimately cooperated. Those who appeared to be less forbidding were re-approached in 77.5% of cases, of whom somewhat more than half cooperated; these are small differences. In addition, the results in table 8.8 suggest that that interviewers in big cities may have been less likely to classify a refusal as absolutely impossible of becoming a future success in the AVO. However, the ESS results in section 10.5 provide indirect evidence that the interviewer' assessment of future cooperation is a useful indicator of reluctance as it correlated with social participation of the respondents: initial refusal who had been characterized as absolutely not willing (and were converted) were on average less politically and socially integrated, had a less favourable attitude towards immigrants and Europe and its institutions, were older and less likely to have high incomes compared to cooperative respondents.

The number of (temporary) refusals or unsuccessful calls appears to be the most objective measure of reluctance. More contacts do not necessarily mean greater reluctance, however, because overcoming reluctance can be achieved within a single contact. This leads to the question of who is more reluctant, a respondent who decided to cooperate after a conversation of half an hour with an interviewer, or someone who just was leaving and said in a hurry 'not interested' but was a pushover at the second contact. A related issue is that experienced interviews might even end up with more 'reluctance' being chalked up in their records because they do not push too much at the first contact and, through a timely retreat, can tailor their approach and maintain contact over contacts so that they come back later to convert a refusal, without exactly making an appointment the first time. Furthermore, when reluctance is overcome across several subsequent contacts with the same or different interviewers, the interaction within subsequent contacts is hardly independent, so adding up contacts might distort what happened.

An additional drawback is that the number of (temporary) refusals may not adequately reflect the degree of reluctance of final refusals. In section 10.5 it was shown that the number of contacts (and the number of temporary refusals) is theoretically a function of the reluctance of the respondent, the decision of the survey organization to re-issue an interview and the probability that an initial refusal is found at home. Therefore, final refusals who refused only once need not be less resistant than final refusals who have refused several times. In the latter case, an interviewer did come back with at least some expectation of obtaining an interview. Furthermore, converted refusals are not necessarily less reluctant than final refusals who were not given the chance to cooperate at a later contact. The nonresponse literature often distinguishes between 'soft' and 'hard' refusals, where soft refusals are those who can be converted and hard refusals are those who cannot. In practice, soft refusals are often those who actually have been converted and hard refusals those who have not been re-approached, re-contacted or converted. Those initial refusals who are not re-approached and re-contacted can usually not be subdivided into soft and hard

refusals and are therefore rather similar to Schrödinger's cat (only when you pop the question (again) do you know if they cooperate or refuse).

In summary, it appears that it is more or less impossible to determine the degree of reluctance of those who did not cooperate at all, and possibly also fairly useless, because there is no survey information on them. The degree of reluctance of those who did cooperate is difficult to ascertain but should always be considered as a result of the fieldwork set-up and the interaction with the interviewer rather than as a feature of the respondent. If possible, the different indicators of reluctance mentioned above should be taken into account. Ideally, information on the doorstep interaction should be collected.

The emphasis on the difficulty of measuring contactability with contact forms and call records makes it easy to forget that the information that was collected in the present study far exceeds the information on fieldwork that is generally available. Without the detailed information from the contact forms the analyses in the previous chapter of contactability, reluctance to cooperate and nonresponse bias could not have been conducted, and the two models presented in section 5.5.4, the continuum of resistance models and the classes of nonparticipants model, could not have been tested. What is more, the overview of the ESS fieldwork in chapter 10 shows that contact forms are an indispensable tool in investigating whether an agreed fieldwork strategy has been adhered to, which problems arose and which improvements can be implemented. And finally, the detailed information from the contact forms (again) provides convincing evidence that contactability and cooperation are different dimensions of survey cooperation (section 7.4) and that those who are hard to contact are equally likely to cooperate as the easy contacts.

11.5 *Do enhanced response rates improve the accuracy of survey results*

Now that it is known which groups are more difficult to contact or less likely to cooperate and why nonresponse should be a reason for concern, high response rates have been obtained and the response process has been measured, the 64-thousand dollar question is whether all these efforts have produced more accurate survey results. This question can be subdivided into three, more focused questions, namely:

- Is there a continuum of resistance with respect to contactability and does enhanced contactability improve survey outcomes?
- Is there a continuum of resistance with respect to cooperation and does converting refusals improve survey outcomes?
- Are there classes of participants similar to classes of nonparticipants and do efforts at including specific classes of participants improve survey outcomes?

In section 5.2 the general continuum of resistance model was described, and section 5.5.4 announced that this would be tested separately for contactability and coopera-

tion. The assumptions with regard to contactability are that potential respondents can be rated on a continuum, that through extended field efforts increasingly hard to contact sample persons can be contacted, that the hardest-to-reach respondents are similar to the final noncontacts and that the accuracy of survey outcomes increases when increasingly hard-to-contact persons are reached. The results from the AVO, presented in section 7.3, provide evidence for the existence of a continuum of resistance with respect to contactability, as there is a linear relationship between the number of calls to first contact and background variables and a number of survey variables. This linear relationship disappears at the end of the continuum for the small number of cases that required a large number of contact attempts, possibly because these had been away from home for extended periods. Section 7.3.3 has also demonstrated that by making more calls and obtaining higher contact rates, the response group becomes more similar to the net sample of the Dutch population, or includes more persons who are generally considered to be hard to reach. This should increase both the precision and the accuracy of survey outcomes (see section 2.3).

Considering the problems with measuring reluctance to cooperate summarized in the previous section, it will not come as a surprise that no clear evidence of the existence of a 'continuum of reluctance' could be found. Section 8.1 showed that it is difficult to conceptualize a continuum of resistance in terms of number of contacts required to obtain cooperation. The evidence presented in section 8.3.1 shows that levels of reluctance do not have a linear relationship with background and survey variables. Furthermore, the results for final refusals in the AVO in section 9.3 indicate that final refusals were quite different from initial refusals, which clearly undermines the idea of a continuum. Worse, enhancing response rates by converting many temporary refusals in the AVO is not likely to have improved the accuracy of outcomes, as the converted refusals were rather similar to the more amenable respondents and somewhat dissimilar from the final refusals. As mentioned above, the composition of the final sample after refusal conversion was even more dissimilar to the population than before.

The ESS results on reluctance in the Netherlands and Germany, however, seem to indicate that temporary refusals differ from immediate cooperators, and hard refusals (more than one refusal conversion attempt) more than soft refusals. These outcomes were found in two countries only, where a substantial number of refusals had been re-approached and converted.

Combining the results for contactability and cooperation in a test of the classes of nonparticipants model, the results are again fairly inconclusive. Classes of participants can be distinguished in the AVO (section 8.3.2) as hard to reach respondents differ from easy respondents and from initially reluctant respondents, although the differences between the last two groups are small. The reluctant participants in the AVO should not be used as a proxy for refusing nonparticipants, however, which

refutes the model. There is some indication, both from the AVO and the ESS, that there is a difference between those who refuse because of situational or of survey-related reasons, and that interviewer impressions of the finality of the refusal may be related to survey characteristics. Due to incomplete evidence on classes of final nonparticipants, the model could not be tested.

11.6 *How to combat nonresponse error and allocate funds effectively?*

Throughout this study it has been maintained that nonresponse is a serious problem, and from the literature and the empirical studies presented, it is clear that nonresponse can cause bias. On the other hand, a number of recent studies presented in section 5.3 and the empirical results from the AVO and the ESS showed that increased fieldwork efforts did not always have the expected effects. The standard solution of a response-enhancer to the threat of nonresponse bias would be to aim for (increasingly) high response rates, as is the strategy of the ESS. This 'hunt for the last respondent' turns out to be increasingly difficult and costly and, in the light of the analyses of refusal conversion described above, may not be altogether effective. The modeller would aim for (increasingly) sophisticated adjustment procedures, which is also difficult because, as summarized in section 11.1, survey cooperation does not have a clear relationship with background variables, or at least not a uniform relationship, and substantive variables that do have this are not generally available for refusals. This is also the reason why a high response rates among a sample of cooperative members of a 'representative' access panel need not reduce nonresponse bias.

The recommendation from a measurer would neither be to settle for low response rates, nor to indiscriminately spend money and deploy extra efforts to recruit extra respondents simply in order to achieve a target response rate when there is no indication that this improves survey quality. What should be done is develop a more scientific approach and create more transparency as regards data collection. Considering the well-developed theory on sampling, measurement and statistical analysis, it seems strange that data collection is often a black hole in between. The conclusion from this study is that the response process of surveys should be closely monitored and, in addition, independent evidence on nonrespondents should be collected to find out whether enhancing response rates does improve accuracy and to have better auxiliary data available to adjust for nonresponse. This could be done through the central question procedure described in section 5.4 or the follow-up survey approach adopted in the present study. An alternative would be to select a random sample of the initial sample before the start of fieldwork and make sure that complete or partial information is collected from a very high percentage of this subsample. In some cases this will be easy, in other cases this will require all the tricks in the book as described in section 6.3. This very high-effort fieldwork collection could be conducted alongside the regular fieldwork among the larger share of the sample

for which an adequate response rate would be pursued. This would solve one of the problems of a follow-up survey, namely the longer duration and possible effects of the period of fieldwork.

The funds required to do this could be raised by settling for a slightly lower response rate for the regular survey than would be feasible. Implementation of this approach may be hazardous, as sponsors do have to be convinced that high target response rates have to be relinquished without a cost reduction. This will be a difficult message to convey to survey users and sponsors as these have come to believe (see Biemer and Lyberg, 2003 (section 2.1)) that high response rate are the outstanding quality feature of surveys. A conceivable risk of telling sponsors that lower response rates are acceptable if a methodological programme is set up that includes monitoring fieldwork progress and outcomes and directly measuring bias, is that they will simply settle for lower response rates and refuse to spend the money saved on methodological 'extras'. This would be a suitable case for refusal conversion.

De Jacht op de Laatste Respondent – Samenvatting

In 1992 schreef de Nederlandse socioloog Herman Vuijsje in opdracht van het ministerie van Onderwijs en Wetenschappen een boekje over de moeizame relatie tussen privacybescherming en wetenschappelijk onderzoek. Daarin bracht hij de paradox ter sprake dat Nederlandse burgers zich niet van de overheid afwenden wanneer deze als subsidieverstrekker optreedt, maar wel als de overheid vraagt om de gegevens die voor de planning van deze voorzieningen nodig zijn. Volgens Vuijsje werd nonrespons bij statistische dataverzameling van overheidswege eerder gerespecteerd en gehonoreerd dan bestreden. Enkele jaren later maakte dezelfde overheid zich alsnog ernstig zorgen over nonrespons, dit naar aanleiding van een boekje van de Nederlandse politicoloog Gerard Visscher (1995), die stelde dat de toenemende nonrespons bij kiezersonderzoek een toenemende desinteresse in de politiek weerspiegelde. De meeste politicologen namen op grond van diezelfde kiezersonderzoeken evenwel aan dat interesse in de politiek stabiel was. Dit leidde tot verhitte discussies in de Nederlandse pers, en in 1996 zelfs tot vragen over nonrespons in de Tweede Kamer. Zeer onlangs stelde de Nederlandse socioloog J.E. Ellemers (2004) in de Sociologische Gids het '*schandaal van de non-respons*' aan de kaak. Dit schandaal houdt in dat er jarenlang stelselmatig is ontkend dat ook bij een nonrespons van 20% of 30% de uitkomsten van een survey-onderzoek dubieus kunnen zijn of dat dit verschijnsel al die tijd is gebagatelliseerd; een nog veel ernstiger zaak vindt hij dat er nauwelijks serieuze aandacht is besteed aan de vraag hoe dit gebrek gerepareerd zou kunnen worden.

De filippica van Ellemers illustreert enerzijds dat nonrespons bij surveys een onverminderd groot probleem is, en anderzijds dat de mogelijk ernstige gevolgen van nonrespons te weinig aandacht krijgen. Dit is des te zorgwekkender omdat het de afgelopen vijftien jaar steeds moeilijker blijkt om een hoge respons te realiseren; mensen zijn minder makkelijk te bereiken en werken minder graag mee aan een onderzoek. Als gevolg hiervan dalen de responspercentages of stijgen de kosten van surveys. Opvallend is hierbij dat Nederland binnen Europa bekend staat om zijn lage responspercentages (zie De Leeuw en De Heer 2002). Omdat surveygegevens een noodzakelijke onderbouwing vormen van publicaties van het Sociaal en Cultureel Planbureau (SCP), omdat de kosten van dataverzameling hoog zijn en men zich in toenemende mate zorgen maakt over mogelijke vertekening die nonrespons kan opleveren, heeft het SCP zich de afgelopen jaren beziggehouden met nonrespons. Deze studie vormt de afronding van dit onderzoek.

Bij surveyonderzoek onder personen of huishoudens wordt een random steekproef uit de bevolking verzocht een vragenlijst te beantwoorden. De vragen hebben meestal vaste antwoordcategorieën en kunnen gaan over feitelijke achtergrondkenmerken van de betreffende persoon of het betreffende huishouden (samenstelling huishouden, leeftijd, type woning, opleiding, betaalde baan), gedrag (museum-

bezoek, stemgedrag, huishoudelijk werk, zorg voor kinderen, contacten met buren) en meningen, houdingen en opinies (vertrouwen in de politiek, motieven bij schoolkeuze, houding ten opzichte van etnische minderheden). Volgens de klassieke steekproeftheorie kunnen de uitkomsten van een random steekproef worden gegeneraliseerd naar de Nederlandse bevolking als geheel. Daarvoor hoeft een steekproef niet heel groot te zijn: 1000 personen is vaak al genoeg. Bij steekproefonderzoek is nonrespons een complicerende factor. Nonrespons treedt op als een deel van de steekproef niet aan het onderzoek meedoet. Dat is vooral vervelend als de respons onder verschillende groepen in de samenleving uiteenloopt. Nonrespons kan leiden tot onjuiste uitkomsten van surveys, en het bestrijden ervan kost tijd en geld. Waarom en wanneer nonrespons tot vertekening leidt, wordt hieronder toegelicht.

In Nederland is een respons van 50% geen uitzondering. Dit betekent dat aan de helft van de oorspronkelijke steekproef wel tijd en geld is besteed, maar dat dit niet resulteert in een ingevulde vragenlijst. De 50% die wel meedoet, vergt daarnaast veel meer inspanning dan voorheen, iets dat de dataverzameling moeizaam, traag en duur maakt. Om de kosten van dataverzameling binnen de perken te houden, kan men ook genoegen nemen met een kleinere steekproef; de resultaten zijn dan echter meestal minder betrouwbaar en er kunnen geen uitspraken worden gedaan over kleine deelgroepen. Een ander probleem is dat de responsgeneigdheid niet bij alle groepen in de Nederlandse samenleving even groot is. Zo zijn bewoners van grote steden notoir moeilijk te enquêteren en kan nonrespons onder deze groep tot een oververtegenwoordiging van dorpelingen en een ondervertegenwoordiging van stedelingen leiden. Dit kan worden voorkomen door in grote steden een extra grote steekproef te trekken, of door met statistische technieken voor een onevenredige vertegenwoordiging te corrigeren, ofwel te wegen. Het grootste probleem treedt op als nonrespons samenhangt met het onderwerp van het onderzoek, en als deze samenhang niet een indirect gevolg is van kenmerken waarvoor gewogen kan worden. Onderstaand voorbeeld kan dit illustreren.

Bij een onderzoek naar gevoelens van onveiligheid kan het gebeuren dat de nonrespons onder bewoners van grote steden groot is, bijvoorbeeld omdat interviewers niet graag 's avonds gaan enquêteren in grootstedelijke achterstandsbuurten, omdat studenten in grote steden vaker op stap gaan en daardoor moeilijk te bereiken zijn, of omdat veel oudere alleenstaande vrouwen in grote steden er niet over peinzen om een vreemde interviewer binnen te laten. Deze nonrespons kan grote gevolgen hebben voor de gerapporteerde gevoelens van onveiligheid. Dit probleem valt niet op te lossen door meer bewoners van grote steden interviewen, als die extra respondenten vooral uit gezinnen met kinderen komen. Meer alleenstaande oudere vrouwen in de grote stad benaderen kan zelfs tot extra vertekening leiden, als de interviewer er alleen in slaagt binnen te komen bij oudere alleenstaande vrouwen die geen last hebben van onveiligheidsgevoelens. Door meer grotestadjongeren te benaderen neemt de vertekening ook niet af als die extra jongeren juist niet tot de

uitgaansjeugd behoren, en dus heel andere ervaringen hebben met veiligheid op straat. Wegen voor nonrespons levert in dit geval weinig op, omdat binnen een bepaalde groep de mensen die niet meedoen verschillen van de mensen die wel meedoen. In deze gevallen kan nonrespons leiden tot onherstelbare vertekening, en dit is waar Ellemers op doelt.

In hoofdstuk 1 en 2 van deze studie wordt uitgebreid uiteengezet waarom nonrespons een probleem is. In hoofdstuk 2 komen daarnaast verschillende soorten steekproeven en verschillende typen surveys aan de orde. De studie is vooral gewijd aan non-respons in face-to-face-onderzoek, dat wil zeggen: surveys waarbij een interviewer respondenten bezoekt en hen vragen stelt. Eventueel kan er nog een schriftelijke vragenlijst worden achtergelaten die de respondent zelf kan invullen. Bij andere vormen van onderzoek wordt gebruikgemaakt van per post verstuurde vragenlijsten, telefonische enquêtes of internetvragenlijsten.

In hoofdstuk 2 volgt een uiteenzetting van de verschillende oorzaken van nonrespons bij surveys. Ten eerste kan het voorkomen dat aan personen in de steekproef nooit wordt gevraagd of ze mee willen werken, omdat ze veel op reis zijn en nooit thuis worden getroffen, de telefoon niet opnemen, post ongelezen weggooien of omdat e-mailtjes niet door een spamfilter komen. Ten tweede kùnnen mensen soms niet meedoen, bijvoorbeeld omdat ze ziek of bedlegerig zijn, verstandelijke beperkingen hebben, niet kunnen lezen of de taal niet begrijpen waarin ze worden geënquêteerd. Ten slotte is het ook mogelijk dat personen in de steekproef die wel zijn bereikt en die wel kunnen meewerken, dat gewoon weigeren: ofwel om algemene redenen (te druk, geen tijd, melk kookt net over), ofwel om meer specifieke redenen (het onderwerp interesseert me niet, ik vertrouw de opdrachtgever niet, ik wil geen contact met onbekenden, ik vind surveys geldverspilling). Uit eerder onderzoek blijkt dat deze drie redenen voor nonrespons, namelijk: onbereikbaarheid, onvermogen en onwil, zorgvuldig uit elkaar gehouden moeten worden en dat het relatief onafhankelijke verschijnselen zijn. Dit houdt in dat mensen die minder makkelijk zijn te bereiken, niet noodzakelijk minder bereid zijn om mee te werken (Lynn et al. 2002). Het onderscheid tussen deze drie oorzaken is van belang omdat de respons kan worden verhoogd door meer mensen te bereiken, deelname voor meer mensen mogelijk te maken (bijvoorbeeld door terug te komen als zieken weer beter zijn of de vragenlijst te vertalen) of meer mensen over te halen. Dat vereist verschillende strategieën. Het onderscheid is tevens van belang omdat nonrespons heel verschillende consequenties kan hebben. Voor een onderzoek naar maatschappelijke participatie kan het immers veel uitmaken of mensen niet meedoen wegens onbereikbaarheid (veel op reis), onvermogen (bedlegerig) of onwil (geen contact willen met onbekenden). In deze studie wordt waar mogelijk een onderscheid gemaakt tussen de verschillende redenen voor nonrespons. ‘Onvermogen’ komt veel minder voor dan ‘onbereikbaarheid’ en ‘onwil’, vandaar dat de meeste aandacht uitgaat naar het onderscheid tussen bereikbaarheid en bereidwilligheid.

Nonresponsonderzoek en centrale vragen

Onderzoekers bestuderen het probleem van de nonrespons op drie verschillende manieren. De eerste groep onder hen richt zich op het verhogen van de respons. Dit kan door een vragenlijst interessant te maken, goede interviewers in te zetten, kleine cadeautjes te geven, et cetera. Bij onderzoek naar technieken om de respons te verhogen (zie § 4.3) staan vragen centraal zoals: welk soort aanschrijfbrieven heeft het meest effect, hoe hoog moet een beloning zijn, moet die vooraf of achteraf worden gegeven, en werkt geld beter dan een schenking aan een goed doel. Daarnaast blijkt een zorgvuldige aanpak erg belangrijk, zoals goede monitoring van het veldwerk, respect voor de respondenten en inzet van gemotiveerde interviewers.

De tweede groep onderzoekers richt zich vooral op het ontwikkelen van modellen van nonrespons en het corrigeren van onderzoeksresultaten met behulp van weegtechnieken. Hierbij beperkt men zich niet tot achtergrondvariabelen, maar probeert men waar mogelijk gegevens te gebruiken die samenhangen met het onderwerp van een survey en met het responsgedrag (zie bijvoorbeeld Bethlehem 2002 en Voogt 2004). Deze studie is opgezet vanuit het standpunt van een derde groep nonresponsonderzoekers, die zicht richt op het meten van (de effecten van) nonrespons en daarmee een middenpositie inneemt tussen de 'verhogers' en de 'corrigeerders'. Vanuit de positie van een 'meter' wordt in deze studie getracht centrale vragen rond nonrespons te beantwoorden, eerst op basis van eerder onderzoek dat in de internationale nonresponsliteratuur is beschreven, en vervolgens aan de hand van eigen onderzoek naar de effecten van nonrespons, uitgevoerd in het kader van het Aanvullend voorzieningengebruik onderzoek (AVO) van het SCP. De inhoud en opzet van dit onderzoek worden beschreven in hoofdstuk 6; daarna komen de analyses van nonrespons aan bod in de hoofdstukken 7 tot en met 9. In hoofdstuk 10 worden ter vergelijking resultaten gepresenteerd van het European Social Survey (ESS), deels omdat in dit onderzoek een hoog responspercentage (70%) wordt nagestreefd, deels omdat bij dit survey een vergelijkbaar meetinstrument wordt gebruikt als bij het AVO, en deels om te toetsen of de conclusies op basis van het AVO ook opgaan voor een onderzoek naar een heel ander onderwerp dat bovendien in ook andere landen is uitgevoerd. De gegevens voor het AVO en het ESS in Nederland werden verzameld door GfK Panel Services Benelux bv.

In paragraaf 1.3 van dit rapport wordt een aantal centrale vragen geïntroduceerd, die in de hoofdstukken 1 tot en met 5 worden besproken aan de hand van de internationale literatuur over nonrespons, en die vervolgens op grond van de analyses van de nonrespons in het AVO en het ESS worden besproken in de hoofdstukken 6 tot en met 10. In deze samenvatting is de scheiding tussen literatuur en empirie opgeheven, en wordt de beantwoording van verschillende vragen gecombineerd. De centrale vragen luiden:

- Waarom moeten we ons zorgen maken over (een hoge) nonrespons?
- Welke personen doen minder vaak mee aan surveyonderzoek omdat ze ofwel moeilijker te bereiken zijn, ofwel minder bereid zijn mee te werken?

– Hoe kan respons worden verhoogd?
– Hoe kan nonrespons worden bestudeerd, dat wil zeggen welke instrumenten en modellen kunnen inzicht geven in nonrespons?
– Leidt een hogere respons tot meer accurate uitkomsten van het onderzoek en tot minder vertekening?
– Hoe kunnen vertekeningen als gevolg van nonrespons het best worden bestreden en hoe kan het onderzoeksbudget hiervoor zo effectief mogelijk worden ingezet?

De eerste vraag kwam al ter sprake in de introductie van deze samenvatting; meer informatie hierover staat in de hoofdstukken 1 en 2. Voor het antwoord op de tweede vraag is vooral geput uit de resultaten van eerder onderzoek. De 'moeilijke' groepen die in de literatuur zijn geïdentificeerd zijn zoveel mogelijk opgenomen in de analyses van het AVO. In deze samenvatting wordt het antwoord op de tweede vraag eerst gegeven voor bereikbaarheid en vervolgens voor bereidwilligheid. Het antwoord op de vraag hoe de respons kan worden verhoogd (door te zorgen voor meer contact en meer medewerking) wordt hier rechtstreeks gekoppeld aan het antwoord op de tweede vraag. De vierde vraag gaat over instrumenten en modellen die de studie van nonrespons mogelijk maken. Met behulp van een aantal instrumenten en modellen is geprobeerd de vijfde vraag te beantwoorden, namelijk of, hoe en in welke mate een hogere respons daadwerkelijk leidt tot meer accurate uitkomsten. Voor de beantwoording van de zesde vraag wordt al het materiaal nog eens in ogenschouw genomen en wordt afgewogen hoe het budget voor dataverzameling zo kan worden ingezet dat de uiteindelijke resultaten optimaal zijn.

Welke personen zijn moeilijk te bereiken en hoe kan de bereikbaarheid worden verhoogd?
Paragraaf 2.6.2 bevat een stapsgewijs overzicht van hoe het contact bij face-to-face-onderzoek tot stand komt en paragraaf 3.2 een opsomming van de slecht te bereiken groepen. Uit eerder onderzoek blijkt dat stedelijkheid en bereikbaarheid negatief correleren. Dit kan verschillende oorzaken hebben. Het kan in stedelijke gebieden moeilijker zijn om interviewers te werven of te behouden, interviewers gaan in sommige stedelijke buurten niet altijd graag 's avonds enquêteren. Dat kan betekenen dat aan stadsbewoners minder energie kan worden besteed (zie ook Bethlehem en Schouten 2003). Daarnaast is het door hoogbouw en intercoms in grote steden moeilijker contact te leggen en hebben bewoners van grote steden minder vaak een bekend telefoonnummer; ook dit maakt het leggen van contact moeilijker. Daarnaast verschilt de bevolkingssamenstelling van grote steden van die van andere gebieden. Jonge alleenstaanden wonen vaker in grote steden en zijn moeilijker te bereiken. Dat laatste is deels een rechtstreeks gevolg van de omvang van het huishouden: hoe meer personen in het gezin, hoe groter de kans dat er iemand thuis is die op zijn minst kan vertellen wanneer de gezochte persoon weer thuis zal zijn, en die in het beste geval gelijk mee kan doen. Resultaten van Nederlands onderzoek naar de tijdsbesteding tonen daarnaast aan dat vrouwen vaker thuis zijn dan mannen, ouderen vaker dan jongeren, en huisvrouwen, werklozen en gepensioneerden vaker dan scholieren,

studenten en werkenden. Een laatste reden waarom mensen minder makkelijk te bereiken zijn is een langdurig verblijf elders. Zo blijkt uit Duits onderzoek dat de hogere nonrespons onder etnische minderheden deels verklaard kan worden uit het feit dat oudere Turken ook buiten de vakantieperiode langere tijd achtereen in het land van herkomst verblijven (Blohm en Diehl 2001).

De resultaten voor bereikbaarheid lijken voor de hand te liggen en geven ook duidelijk aan hoe nonrespons door het uitblijven van contact kan worden verminderd. In hoofdstuk 7 wordt nader ingegaan op de wijze waarop bij het AVO contact werd gezocht met de huishoudens in de steekproef; in dat onderzoek werden op een paar procent na alle huishoudens uit de steekproef bereikt. Uit de analyse blijkt dat interviewers bij voorkeur probeerden overdag contact te maken, terwijl de kans op contact 's avonds groter is. In latere fasen van het veldwerk mochten slecht bereikbare huishoudens telefonisch worden benaderd. Veel contactpogingen op verschillende tijden van de dag en op verschillende dagen van de week, en telefonische benadering van mensen die vaak niet thuis zijn, blijkt tot een zeer hoog contactpercentage te kunnen leiden. Daarnaast is het nuttig om niet alle contactpogingen in een te korte periode te laten vallen, om ook mensen die wat langer van huis zijn te kunnen aantreffen.

In de analyse van bereikbaarheid bij het AVO werden grotendeels dezelfde slecht bereikbare groepen geïdentificeerd als in de literatuur. Daarnaast blijkt dat mensen met een bovengemiddelde culturele participatie (bezoek aan concerten, popconcerten, film, dance events, ballet, et cetera) extra moeilijk te bereiken zijn. Omdat culturele participatie een van de hoofdonderwerpen van het AVO is, is het hoge percentage huishoudens dat bereikt is, dus van groot belang.

Welke personen zijn minder bereidwillig of weigeren mee te werken en hoe kan de bereidwilligheid worden verhoogd?

De in hoofdstuk 3 weergegeven bevindingen voor bereidwilligheid zijn veel gecompliceerder dan die voor bereikbaarheid. Eerder onderzoek geeft aan dat het moeilijk of zelfs onmogelijk is om op eenduidige wijze de achtergrondvariabelen te identificeren die de kans op medewerking bepalen en dat de groep mensen die nooit meedoet waarschijnlijk heel klein is (Goyder 1987; Elliot 1991; Schnell 1997). Vaak worden in de analyse van nonrespons verschillende leeftijdsgroepen onderscheiden. Leeftijd op zich is echter geen verklarende factor. Ouderen kunnen weigeren omdat ze geen vreemden in huis willen laten, omdat ze denken dat ze voor de onderzoeker niet interessant zijn omdat ze veel minder werken, reizen of uitgaan, of omdat ze terugschrikken voor een gecompliceerde vragenlijst. Anderzijds kan hun medewerking weer hoger liggen omdat ouderen meer plichtsgevoel hebben of meer gezagsgetrouw zijn. Groves en Couper (1998) vonden uiteindelijk geen eenduidig effect van leeftijd op deelname aan surveys.

Een additioneel probleem is dat achtergrondkenmerken vaak onderling samenhangen, waardoor het moeilijk is de bepalende factor te identificeren. Onder ouderen zijn bijvoorbeeld relatief veel alleenstaande vrouwen met een lagere opleiding. Nog moeilijker is het om precies te achterhalen waarom de respons in de grote steden vaak laag is. Komt dat door de afwijkende bevolkingssamenstelling in de grote steden (meer lage inkomens en meer niet-westerse allochtonen), door de zwakkere sociale cohesie en sterkere gevoelens van onveiligheid in de grote stad, of door een combinatie hiervan? Iets vergelijkbaars geldt voor de lagere respons onder personen uit etnische minderheidsgroepen. Hangt lagere bereidwilligheid onder leden van deze groepen samen met het grotestadeffect, met taalproblemen, een lagere opleiding, terughoudendheid om een mannelijke interviewer binnen te laten dan wel een vrouwelijke interviewer te woord te staan? Ook hier moet men ervoor waken correlaten van een lagere responsgeneigdheid te verwarren met de oorzaken.

In hoofdstuk 4 is daarom dieper ingegaan op de achterliggende redenen van deelname aan surveys. Paragraaf 4.2 bevat een overzicht van eerder onderzoek naar de relatie tussen responsgedrag, sociale isolatie en maatschappelijke betrokkenheid, en van eerder onderzoek waarin respondenten rechtstreeks naar hun oordeel over surveys zijn gevraagd. Uit de literatuur blijkt dat medewerking aan surveys samenhangt met een aantal factoren die niet of niet volledig met achtergrondkenmerken verklaard kunnen: sociale isolatie en sociale participatie, zich houden aan de wet, maatschappelijke betrokkenheid, het verrichten van vrijwilligerswerk, kennis van en interesse in de politiek, politiek cynisme, deelname aan verkiezingen en ervaring met het invullen van formuleren. Daarnaast hangt nonrespons samen met leefstijl, deel uitmaken van de dominante cultuur, pc-bezit en internetgebruik, en type vrijetijdsbesteding. Ten slotte is het meetinstrument (mondeling, telefonisch, schriftelijk onderzoek) zelf ook een bepalende factor en is deelname afhankelijk van het type veldwerkorganisatie en opdrachtgever (overheid, universiteit, commerciële instelling), de houding tegenover surveys (zijn surveys nuttig, interessant, leuk) en de interesse in en betrokkenheid bij het onderwerp van een survey. Uit paragraaf 3.3.5 blijkt dat drukke bezigheden niet per se een reden zijn om niet mee te doen. Er zijn zelfs aanwijzingen dat drukke mensen net zo makkelijk ook nog even een vragenlijst invullen.

Uit bovenstaand overzicht blijkt dat nonrespons vertekening op kan leveren. Anderzijds bevat het overzicht ook aanwijzingen voor het verhogen van de respons. Ten eerste moet een goed onderzoek voldoen aan een aantal algemene regels, zoals het inzetten van goede en ervaren interviewers, het verstrekken van adequate informatie aan respondenten (door middel van een aanschrijfbrief, folder en eventueel een website, maar ook via de interviewer of via een telefonische helpdesk) en het ontwikkelen van een duidelijke vragenlijst met heldere vragen. Zoals hierboven vermeld is er veel onderzoek gedaan naar de meest effectieve inhoud van aanschrijfbrieven en de inzet van *incentives* (kleine cadeautjes voor de respondenten). Ten tweede kunnen er veel verschillende redenen zijn waarom mensen niet mee willen doen. In plaats van

een standaard aanpak komt het er dan op aan dat de interviewer over een breed scala van argumenten en beloningen beschikt die ieder een bepaald type bezwaar kunnen wegnemen. Groves en McGonagle (2001) noemen het handhaven van de interactie tussen interviewer en respondent en het bij die interactie inspelen op de kenmerken, voorkeuren en interesses van de individuele respondent als bepalende factoren. Vroeger konden mensen het nog wel als een eer beschouwen dat naar hun mening werd gevraagd, en hoefde de interviewer weinig overtuigingskracht te hebben. Nu komt het er op aan dat de interviewer goed en snel inschat welke redenen om mee te doen voor welke respondent het zwaarst wegen, en de juiste argumenten bij de hand heeft om de respondent over te halen: het is een belangrijk onderwerp, we zijn geïnteresseerd in uw mening, de resultaten worden gebruikt voor ontwikkeling van beleid, voor deelname krijgt u een leuk cadeautje, de resultaten komen in de krant, het kan op een tijdstip dat het u schikt, nu kunt u eindelijk eens uw kritiek spuien, dit onderzoek is van groot wetenschappelijk belang, et cetera, et cetera.

Strategieën om deelname te verhogen worden uitgebreid beschreven in paragraaf 4.3. In hoofdstuk 6 wordt uiteengezet hoe deze kennis is toegepast in de praktijk, en staat uitgebreid beschreven hoe in het AVO, een onderzoek met lange vragenlijsten voor alle leden van het huishouden, een hoge respons van 65% is bereikt. Hier komt ook het vervolgonderzoek aan de orde onder degenen die pertinent weigerden deel te nemen aan het AVO. Hierbij werd een respons van 70% gerealiseerd. In paragraaf 6.3 is uiteengezet waarom dit vervolgonderzoek een succes was en wordt de effectiviteit van verschillende responsverhogende middelen geïllustreerd. In hoofdstuk 10 staat in het kort welke responsverhogende strategieën in het European Social Survey (ESS) zijn toegepast en welk effect dit had op de respons. Bij dit onderzoek naar waarden, attituden, meningen en gerelateerd gedrag, waarvoor in 2002/2003 voor het eerst data werden verzameld, werd een responspercentage van 70% nagestreefd, hetgeen in niet alle deelnemende landen werd gerealiseerd. Uit figuur 10.1 blijkt dat Nederland heel dicht in de buurt kwam van het streefpercentage.

In de praktijk blijken de mogelijkheden om minder bereidwillige respondenten over de streep te halen, meestal beperkt. Gelukkig is er sprake van een tweede kans, als eerdere weigeraars nog eens benaderd worden. In de literatuur wordt dit *refusal conversion* genoemd (zie §2.6.3). Heel wat mensen blijken alsnog ja te zeggen als er een andere interviewer aan de deur komt, of als dezelfde interviewer nog een keer langskomt op een mogelijk geschikter moment. In hoofdstuk 8 wordt duidelijk dat de respons op het AVO aanzienlijk is verhoogd dankzij herbenadering van weigeraars. Hetzelfde geldt voor het ESS in Nederland, maar uit hoofdstuk 10 blijkt ook dat 'refusal conversion' in veel andere Europese landen aanzienlijk minder vaak werd toegepast of aanzienlijk minder vaak succes had.

Bij het herbenaderen van weigeraars ligt het voor de hand dat men alleen teruggaat naar veelbelovende 'zachte' weigeraars. Dat maakt het herbenaderen makkelijker,

maar kan betekenen dat de overgehaalde weigeraars een aselecte subgroep van alle weigeraars vormen. Daarnaast is de inschatting of een weigeraar misschien later wel meedoet, afkomstig van een interviewer die geen respons heeft gekregen en die daarmee niet altijd de beste beoordelaar van toekomstig succes hoeft te zijn, zoals ook blijkt uit hoofdstuk 8.

Uit de analyse van het AVO in de hoofdstukken 8 en 9 blijkt eens te meer dat achtergrondkenmerken in de analyse van bereidwilligheid om mee te doen een moeilijk te interpreteren rol spelen. Bewoners van grote steden waren in eerste instantie bijvoorbeeld minder bereidwillig om aan dit onderzoek mee te doen, maar werden uiteindelijk in groten getale overgehaald. Alleenstaande mannen deden in eerste instantie weinig mee en lieten zich in tweede instantie nog veel minder vaak vermurwen om toch mee te doen. Dit had als gevolg dat onder de bereidwillige respondenten grotestadbewoners en mannelijke alleenstaanden waren ondervertegenwoordigd, dat er onder overgehaalde respondenten veel grotestadbewoners en weinig alleenstaande mannen waren, en dat de uiteindelijke groep non-respondenten relatief weinig grotestadbewoners en relatief veel alleenstaande mannen telde.

Uit de hoofdstukken 8 en 9 blijkt verder dat de verschillen tussen makkelijke respondenten, aanvankelijk weigerende respondenten (de overgehaalde weigeraars) en uiteindelijke weigeraars (die wel aan het vervolgonderzoek deelnamen) in het AVO relatief klein waren. Overgehaalde weigeraars lijken nog het meest op de doorsnee burger en beantwoorden zeker niet aan het stereotype van personen die minder sociaal participeren. De uiteindelijke weigeraars beantwoorden iets meer aan dat beeld. Die doen minder aan sport, maar gaan vaker naar popconcerten, dance events en films; ook hebben ze minder vaak een pc, maar als ze die hebben dan maken ze vaker gebruik van internet en gebruiken ze dit minder vaak voor serieuze toepassingen; verder lezen ze wat minder en behoren ze minder vaak tot een kerkgenootschap.

Overigens blijken de overgehaalde respondenten van het ESS in Nederland wel te verschillen van de eerder bereidwillige: ze zijn minder geïnteresseerd in de politiek en hebben minder vertrouwen in politici. Een mogelijke verklaring voor het verschil met het AVO is dat de onderwerpen in ESS en AVO verschillen. Er is eerder al een verband gelegd tussen surveyparticipatie, maatschappelijke participatie en politieke participatie en interesse. Gezien de inhoud van het ESS zou het onderwerp van dit onderzoek sterker kunnen samenhangen met responsgeneigdheid dan dit het geval is bij het AVO, waarbij een breed spectrum van feitelijke vragen wordt gesteld.

Instrumenten en modellen en het effect van een hogere respons

In het voorgaande is uiteengezet waarom nonrespons een probleem kan zijn, welke kenmerken met nonrespons samenhangen en waarom dat tot vertekening kan leiden, en hoe respons kan worden verhoogd. Om aan te kunnen geven of en hoe een hogere respons tot betere resultaten leidt, en welke vertekening er overblijft,

zijn modellen en meetinstrumenten nodig. Lin en Schaeffer introduceerden in 1995 twee modellen die in veel nonresponsonderzoek impliciet en expliciet worden gehanteerd (zie § 5.2). Het 'weerstandscontinuüm-model' impliceert dat personen op een eendimensionale schaal geplaatst kunnen worden, geordend naar responsgeneigdheid. Hoe intensiever het veldwerk, hoe meer mensen meedoen. De mensen die de meeste inspanning vereisen lijken volgens het model het meest op de uiteindelijke nonrespondenten, en hoe hoger de respons, hoe beter de resultaten. Het 'non-participanten-klassenmodel' gaat ervan uit dat mensen om verschillende redenen meedoen, en dat verschillende typen non-participanten lijken op verschillende typen respondenten: weigeraars bijvoorbeeld op mensen die eerst nee zeggen en dan toch meedoen, en niet-bereikten op moeilijk bereikbaren. Als dit model opgaat, hoeft een hogere respons niet noodzakelijkerwijs tot minder vertekening te leiden. Voor beide modellen geldt dat er voor de toetsing zowel informatie nodig is waarmee men respondenten kan onderscheiden (naar type of naar mate van weerstand), als informatie over nonrespondenten.

Om 'makkelijke' en 'moeilijke' respondenten, of verschillende typen respondenten te kunnen onderscheiden is meer gedetailleerde informatie over het verloop van het veldwerk nodig. Waarom dat zo belangrijk is, en waarom deze informatie zo zelden beschikbaar is, staat beschreven in paragraaf 2.6.4. In hoofdstuk 6 is het contactformulier van het AVO beschreven, het meetinstrument waarmee de benodigde informatie kan worden geregistreerd. Op grond van de resultaten van het huidige onderzoek is dit contactformulier overigens al verbeterd. In hoofdstuk 10 wordt uitgebreid aandacht besteed aan de contactformulieren van het ESS. Een van de belangrijke aanbevelingen van de huidige studie is bij surveyonderzoek het verloop van het veldwerk op dergelijke contactformulieren vast te leggen. Ten eerste kan daarmee worden gecontroleerd of het veldwerk volgens de specificaties is uitgevoerd, en ten tweede kunnen de veldwerkinspanningen en de oorzaken van nonrespons zo precies worden beschreven. Dat laatste is geen doel op zich, maar speelt een rol bij onderzoek naar de relatie tussen veldwerkinspanningen, respons en nonresponsvertekening, en bij een effectievere opzet van het veldwerk. Idealiter zouden de contactformulieren zelfs kunnen dienen om het lopende veldwerk bij te sturen.

Als respondenten die meer of minder veldwerkinspanningen vereisen weinig van elkaar verschillen, dragen die extra, vaak kostbare inspanningen niet bij tot vermindering van de vertekening als gevolg van nonrespons . In paragraaf 5.3 wordt een overzicht gegeven van onderzoek naar het effect van extra veldwerkinspanningen. In recente Amerikaanse studies kwamen Curtin et al. (2000), Keeter et al. (2002) en Teitler et al. (2003) op grond van hun analyses tot de conclusie dat een hogere respons, na correctie voor achtergrondkenmerken, weinig aan de resultaten veranderde. Geen van de auteurs pleit er overigens voor om dan maar met een lage respons akkoord te gaan, al is het maar omdat er ook in hun onderzoeken een substantieel deel weigeraars overbleef dat aanzienlijke vertekening kon veroorzaken. In andere

onderzoeken (bijv. Voogt 2004) blijken respondenten die extra veldwerkinspanningen vereisen wel onderling te verschillen.

De uitspraken die hierboven zijn gedaan over moeilijker te bereiken en minder bereidwillige respondenten in het AVO resp. ESS zijn gebaseerd op de informatie uit de contactformulieren. De resultaten uit het AVO suggereren dat slecht bereikbare respondenten verschillen van makkelijk bereikbare, maar dat meer en minder bereidwilligen weinig onderling verschillen. De resultaten van het ESS wijzen daarentegen wel op een verschil tussen meer en minder bereidwilligen. Hieruit blijkt het nut van een analyse van het effect van extra veldwerkinspanningen met behulp van de contactformulieren, maar ook dat het niet mogelijk is om over verschillende typen onderzoek te generaliseren.

Naast informatie om verschillende typen respondenten te kunnen onderscheiden, vereisen bovengenoemde modellen ook dat er informatie beschikbaar is over de mensen die niet meedoen. Hiervoor kunnen verschillende bronnen worden aangesproken. Ten eerste is het mogelijk dat van alle steekproefpersonen al de nodige informatie bekend is, bijvoorbeeld als de steekproef is getrokken uit een bestand met gegevens uit bestaande registraties. Bij veel steekproeven van het Centraal Bureau voor de Statistiek is dat het geval. Ook kunnen aan de steekproef buurtkenmerken of zelfs het adres worden gekoppeld. Daarnaast kunnen interviewers kenmerken van buurt en woning vastleggen op het contactformulier. De bovengenoemde gegevens over het verloop van het veldwerk zijn natuurlijk ook bekend voor nonrespondenten. En in speciale gevallen is er al eerder contact geweest met de huidige nonrespondenten, bijvoorbeeld in het kader van een panelonderzoek.

Al deze gegevens kunnen helpen bij het schatten van de vertekening, maar als de nonrespons samenhangt met de inhoud van het survey dan schieten ze toch tekort. Paradoxaal genoeg is het effect van nonrespons op surveys pas in te schatten zodra er surveyinformatie van nonrespondenten bekend is. Hiertoe zijn twee instrumenten beschikbaar die in paragraaf 5.4 zijn beschreven. In de 'centrale vraag'-procedure wordt geprobeerd aan alle weigeraars een of enkele vragen te stellen die de kern van het onderzoek raken. In deze studie is het tweede instrument toegepast, namelijk een vervolgonderzoek onder een steekproef van hardnekkige weigeraars. Met de resultaten van dit onderzoek is het mogelijk een goede schatting te maken van de vertekening die nonrespons in het AVO opleverde, omdat de respons bij dit vervolgonderzoek hoog was. De kenmerken van de uiteindelijke weigeraars bij het AVO, die wel alsnog meededen aan het vervolgonderzoek, zijn hierboven beschreven.

Uit de analyses in de hoofdstukken 7, 8 en 9 blijkt dat er bij het AVO sprake is van een weerstandscontinuüm wat betreft bereikbaarheid, hoewel de slechtst bereikbare mensen waarschijnlijk niet degenen waren die veel bezigheden buitenshuis hadden, maar degenen die langere tijd afwezig waren. Op grond van de beschikbare gegevens

is het echter niet mogelijk om een eendimensionale schaal te onderscheiden van oplopende onwil om mee te doen aan het AVO. Dit wordt waarschijnlijk veroorzaakt door het feit dat enerzijds de mate van benodigde veldwerkinspanningen wordt gebruikt om bereidwilligheid te indiceren, en dat anderzijds de veldwerkinspanningen niet worden gericht op mensen waarvan kan worden aangenomen dat ze toch niet mee zullen doen. Ook andere maten van bereidwilligheid zijn echter niet eenduidig te interpreteren.

Het model van toenemende weerstand gaat dus maar zeer ten dele op in het AVO, namelijk wel met betrekking tot bereikbaarheid maar niet met betrekking tot bereidwilligheid. Het model van klassen non-participanten gaat niet op omdat, zoals hierboven al werd beschreven, de overgehaalde weigeraars niet lijken op de uiteindelijke weigeraars. Als gevolg hiervan is het onaannemelijk dat het verhogen van de respons door aanvankelijke weigeraars bij het AVO over te halen, geleid heeft tot een betere kwaliteit van de uitkomsten van het onderzoek en tot een kleinere vertekening als gevolg van nonrespons.

Het einde van de jacht

Uit deze studie is gebleken dat het steeds moeilijker wordt om een hoge respons te bereiken, maar dat dit met de nodige inspanningen wel degelijk is te realiseren. Nederland hoeft zich gezien de resultaten van het Aanvullend voorzieningengebruik onderzoek en het European Social Survey zeker niet te generen. Toch is het niet eenvoudig vertekening als gevolg van nonrespons te bestrijden. Correctie door het wegen met achtergrondkenmerken biedt geen oplossing als nonrespons met het onderwerp van een onderzoek samenhangt, en het streven naar een responspercentage van 70%, dat allerwegen als goed wordt gekwalificeerd, leidt niet noodzakelijkerwijs tot minder vertekening.

In de literatuur wordt er steeds meer voor gepleit om inzicht in de oorzaken van nonrespons te verwerven en meer aandacht te besteden aan de vertekening als gevolg van nonrespons dan aan de hoogte van nonresponspercentages. Daarnaast is het van belang nonresponsvertekening niet geïsoleerd te beschouwen, maar als onderdeel van de totale surveykwaliteit (zie Biemer en Lyberg 2003). Met deze studie is getracht aan dit inzicht bij te dragen met als achterliggend doel het budget voor veldwerk zodanig in te zetten dat de uitkomsten van een survey zo goed mogelijk de situatie van de populatie weergeven, met zo min mogelijk vertekening als gevolg van nonrespons. Het vastleggen van het responsproces op contactformulieren en het verzamelen van informatie over nonrespondenten kan meer bijdragen aan de kwaliteit van surveys dan een verbeten jacht op de laatste respondent.

Annexes

Annex 6.1 Multiple households in the AVO1999

In the AVO1999 up to four households were to be approached per address, if the address comprised multiple households. Selection criteria were not given. The presence of multiple households at an address can sometimes be assessed by the presence of several doorbells next to a single letterbox. In other cases, contact is required and the initial cooperation from a resident who has to be willing to tell the interviewer how many households live at the address. At a multiple household address the interviewer had to fill in a contact form for each household up to a maximum of four. The actual registration of multiple households was rather incomplete and somewhat inconsistent, as shown in table A6.1. Each household at the same address should report the same number of households at this address, for instance. The first row shows that at six addresses two contact forms were used but no contact form mentioned more than one household at the address.[1] In other cases too, the contact forms for different households at the same address provide conflicting information on the number of households at that address. Furthermore, it rarely occurred that contact forms were available for all (or four) households at a given address. A sad case is an address with six households according to the single contact form of the one recorded, nonresponding household. It is not clear whether multiple households have frequently just not been contacted or whether the inconsistencies are simply due to clerical errors. As the number of multiple household addresses is very small (about 0.5%) this is a minor problem. Given the small number of addresses with multiple households in the data file (49) and the even smaller number of addresses with at least one responding multiple household (43), the total number of multiple households (68) will be treated as single households, as only 25 responding households are the second or third at their address.

The deficient registration of multiple households makes it impossible to compute an exact response rate at household level. What is more worrying, however, is the small number of multiple households in the data file. Even in university cities, where shared accommodation for students is common, multiple households are extremely scarce. As mentioned above, only 0.5% of the eligible addresses in the AVO1999 appear to comprise multiple households. According to estimates from Statistics

1 Further inspection showed that at one of these addresses, one form stated that no eligible household resided there and the other mentioned a single, responding household at this address. At four addresses, both contact forms mentioned that one household resided there, one responding and one not responding. And finally, at one address two responding households lived, both with a contact form that stated that one household resided at this address.

Netherlands (GBA, 1 January 2000) the percentage of two-household addresses is 2.1% and of three or more households 1.1%. The difference may be partly due to differences in definition. However, the fact remains that multiple households are likely to be underrepresented in the net AVO sample, which may have a detrimental effect on the socio-demographics of respondents and nonrespondents in bigger cities.

Table 6.1 ddresses with multiple households according to number of households per address (rows) and contact forms (columns)[a]

maximum reported number of households per address	1	2	3	4	N addresses	multiple households
1		6			6	6
2	8	13	1		22	44
3	7	1	5		13	39
4	2	1		2	5	20
5	1			1	2	8
6	1				1	4
N addresses	19	21	6	3	49	121
N contact forms	19	42	18	12	91	

a The total number of contact forms in the data (91) includes three contact forms reporting that no households live at this address (despite the presence of contact forms of households at this address). This leaves a total of 88 households from multi-household addresses in the file.

Annex 6.2 Fieldwork waves in the AVO1999

The fieldwork of the AVO1999 started in September with a first wave in which all addresses were put into the field. In each wave a maximum of four calls could be made. If the first wave did not result in an interview (or an appointment) the questionnaires were returned to the fieldwork agency for reissue. In this second wave, new contact forms were used, and again four calls could be made. After this second wave, the successful sample addresses were again put into the field for a third wave.

Table A6.2 presents the number of calls by call number per wave.[1] The table shows that the first call to each address was always made in the first wave (9394). In this wave 5771 households received a second call, 3230 a third and 1161 a fourth. It also occurred, however, that the second call for a household only took place in the second wave (1960). The households received a call in wave 1, were then left alone and reapproached only after reissue of the interviews in wave 2. In each wave only four calls could be made. In the first wave this could only be the first four calls. In wave 2 this could be call 2 to 8, depending on what happened in the first wave. The 106 households that received an eighth call in wave 2 must have been called on four times in wave 1 and four times in wave 2. The total number of calls in wave 1 was 19,556, in wave 2 9,968 and in wave 3 3,878.

In the columns under Waves (calls to first contact) those calls have been left out that were made to previously contacted respondents. If contact was established at the second call (appointment) and the interview was conducted at the third call, this third call is not included in the right-hand part of the table. The left-hand part of the table is the best indicator of fieldwork load across waves, the right part the best indicator of contactability across waves. The bottom row shows that 12% of all calls took place in the third wave, but that only 2% of all calls to previously noncontacted households took place in the third wave. This final wave may therefore have been more effective for refusal conversion purposes than for increasing the contact rate.

1 As always a call is a contact attempt that may be successful (contact with respondent) or unsuccessful (nobody at home). A contact may be successful (interview) or unsuccessful (refusal) or a preliminary to success (an appointment). An undocumented share of the later calls per wave may have been telephone calls.

Table 6.2 ieldwork waves (columns) and calls (rows)

	waves (all calls)				waves (calls to first contact)			
	1	2	3	total	1	2	3	total
1	9,394			9,394	9,394			9,394
2	5,771	1,960		7,731	4,677	195		4,872
3	3,230	2,244	371	5,845	2,514	254	4	2,772
4	1,161	2,357	637	4,155	838	585	9	1,432
5		1,974	733	2,707		677	24	701
6		961	780	1,741		365	59	424
7		366	613	979		156	89	245
8		106	401	507		51	100	151
9			241	241			92	92
10			78	78			39	39
11			21	21			8	8
12			3	3				0
total	19,556	9,968	3,878	33,402	17,423	2,283	424	20,130
% wave	59	30	12	100	87	11	2	100

wave 1: September-November 1999; wave 2: October 1999-January 2000; wave 3: November 1999-February 2000

Annex 6.3 Interviewer assignments in the AVO1999

For each call an interviewer identification had to be available. However, about 2000 of the 10,000 initial calls at households in the Randstad (the western and most densely populated part of the Netherlands containing the three largest cities) were conducted by a subcontractor. This occurred also in subsequent waves, although less often (400 cases in wave 2, 100 in wave 1). Thus assignment sizes could not be computed for all interviewers. For the majority of cases where interviewers were identifiable, assignment sizes and deployment over waves have been computed (see table A6.3). In the first wave, interviewers were assigned approximately 39 cases each. Across waves, interviewers were assigned approximately 66 cases. As the table shows, there are large differences in interviewer assignments. Most interviewers, about 81%, were deployed in all three waves, 6% only in the first wave(s) and 13% only in later wave(s).

Table 6.3 Interviewer assignment sizes in first wave, sum of wave 1, 2 and 3 and interviewer deployment across waves

	wave 1		wave 1, 2, 3	
assignment size	n	%	n	%
1-10	36	18	26	12
11-20	28	14	19	8
21-30	26	13	18	8
31-40	26	13	21	9
41-50	21	11	16	7
51-60	23	12	15	7
61-70	11	6	20	9
71-80	8	4	22	10
81-90	6	3	12	5
91-100	4	2	11	5
101-110	1	1	14	6
111-120	4	2	6	3
121-130	1	1	6	3
131-140			6	3
141-150				0
151-160			2	1
161-170			1	0
171-180	1	1	2	1
181-190			3	1
191-200	1	1	1	0
more than 200			5	2
total	197		226	
average	39		66	
deployment interviewers	n	%		
wave 1	9	4		
wave 1 + 2	4	2		
wave 1 + 2 + 3	184	81		
wave 2 + 3	17	8		
wave 3	12	5		
total	226	0		

Annex 6.4 Contact form AVO1999

KRT.01

ADDRESS ACCOUNT WAVE 1

		FACE-TO-FACE			*TELEPHONE*	
		Call 1	**Call 2**	**Call 3**	**Call 4**	
DATE	1 Monday	.....	.....	.....	.....	11-14
	2 Tuesday	.....	.....	.....	.....	
	3 Wednesday	.....	.....	.....	.....	
	4 T ursday	.....	.....	.....	.....	
	5 Friday	.....	.....	.....	.....	
	6 Saturday	.....	.....	.....	.....	
DAIL PRIOD	Morning	-1	-1	-1	-1	15-18
	Afternoon	-2	-2	-2	-2	
	Evening	-3	-3	-3	-3	
MONTH	September 1999	-09	-09	-09	-09	19-26
	October 1999	-10	-10	-10	-10	
	November 1999	-11	-11	-11	-11	
	December 1999	-12	-12	-12	-12	
	January 2000	-01	-01	-01	-01	

3a. **IF CONTACT WAS MADE B TELEPHONE:**
How many times did you call before t e appointment was arranged?

[ ] **times** 27-28

3b. Did you get an answering mac ine or voicemail telling you to leave a message?

Yes, answering mac ine or voicemail .....-1
No, no answering mac ine or voicemail .....-2 29

IF CONTACT WAS MADE FACE-TO-FACE:
3c. Was t ere an entryp one or did you speak personally (face-to-face) wit t e respondent?

Entryp one only.....-1
Face-to-face only.....-2
First entryp one, t en face-to-face .....-3 30

(8-99-8596-Oms/v2)

ADDRESS ACCOUNT WAVE 1

RESULT:	Call 1	Call 2	Call 3	Call 4		
Interview	-01	-01	-01	-01		31-32
Appointment at (DATE)	-02	-02	-02	-02		
Interview broken off	-03	-03	-03	-03	→ GO TO Q. 4a	33-34
Not at ome	-04	-04	-04	-04		
REASON NO INTERVIEW:						
Vacant adress	-05	-05	-05	-05		
Probably at ome, door not opened	-06	-06	-06	-06	→ GO TO Q. 4b	35-36
Not able, due to illness	-07	-07	-07	-07	→ GO TO Q. 4c	
Not able because of activities, visitors	-08	-08	-08	-08	→ GO TO Q. 4d	37-38
Not able because of language problems **INT.: RECORD LANGUAGE + COUNTR OF ORIGININ**	-09	-09	-09	-09	→ GO TO Q. 6	
Language :	.....	.....	.....	.....		39-40
Country of origin :	.....	.....	.....	.....		41-42
Not able because of ot er reasons	-10	-10	-10	-10	→ GO TO Q. 4e	
Does not want to participate: refusal, did not feel like cooperating	-11	-11	-11	-11	PROCEED WITH Q. 5	
Does not want to participate: ot er reason	-12	-12	-12	-12		

INT.: DO NOT FORGET NEIGHBOURHOOD CHARACTERISTICS ON REVERSE!

4a. **BROKEN-OFF INTERVIEW**: W y was it broken off? 43-44 45-46 47-48

(W en) can t e interview be completed? 49-50 51-52 53-54

Status: W ic questionnaires ave been completed?

Completed		Not completed		
House old form	-1	House old form	-2	55
Individual form (Adults)	-3	Individual form (Adults)	-4	56-57
Yout	-5	Yout	-6	58-59

→ **PROCEED WITH QUESTION 6**

4b. **PROBABL AT HOME, BUT … :** On t e basis of w at did you record t is reason? 60-61 62-63 64-65

→ **PROCEED WITH QUESTION 6**

4c. **NOT ABLE DUE TO ILLNESS:** W en will e or s e ave recovered? (possibly new appointment) 66-67 68-69 70-71

→ **PROCEED WITH QUESTION 6**

(8-99-8596-Oms/v2)

ADDRESS ACCOUNT WAVE 1

4d. **NOT ABLE BECAUSE OF ACTIVITIES:**
When can a new appointment be made?

72-73
74-75
76-77

→ **PROCEED WITH QUESTION 6**

4e. **NOT ABLE BECAUSE OF OTHER CIRCUMSTANCES:**

8-9
10-11
12-13

→ **PROCEED WITH QUESTION 6**

5. **INT.: PLEASE INDICATE IN AS MUCH DETAIL AS POSSIBLE WH THE RESPONDENT DOES NOT WANT TO COOPERATE (AT THIS MOMENT).**

14-15
16-17
18-19

INT: GIVE OUR OWN ASSESSMENT.

RESPONDENT :

20

- Absolutely does not want to be re-approached ..... -1
- Would rather not be re-approached ..... -2
- Is willing to participate, but only if paper-and-pencil ..... -3
- Is willing to participate, but only if telephone ..... -4
- Will possibly cooperate after some time ..... -5
- Will probably cooperate after some time ..... -6
- Don't know ..... -7

INT.: ALWA S FILL IN OVERLEAF (NEIGHBOURHOOD CHARACTERISTICS)!

(8-99-8596-Oms/v2)

Annex 6.5 Interviewer observation form AVO1999

KRT.04

6. **INT.: PLEASE RECORD SEX AND AGE OF RESPONDENT** 44

Female ..... -1
Male ..... -2

AGE (ESTIMATE IF NECESSAR):

[ ] **ears** 45-46

INFORMATION ON DWELLING AND NEIGHBOURHOOD

INT.: THESE ITEMS ARE TO BE ANSWERED B THE INTERVIEWER FOR BOTH RESPONDENTS AND NONRESPONDENTS!

7a. Is t e neig bour ood in w ic t e dwelling is situated: 47

A 'pre-war' neig bour ood (built before 1945) ..... -1
A more recent neig bour ood (predominantly between 1946 and 1974) ..... -2
A neig bour ood predominantly built after 1974 ..... -3

7b. Is t is a neig bour ood comprising: 48

Predominantly low-rise buildings ..... -1
Predominantly ig -rise buildings ..... -2
As many ig -rise as low-rise buildings ..... -3

7c. How would you describe t e maintenance of t e *neighbourhood*: 49

Excellent ..... -1
Good ..... -2
Mediocre ..... -3
Poor ..... -4

7d. Is t e neig bour ood located: 50

In t e centre ..... -1
Rig t next to t e centre ..... -2
Outside t e centre ..... -3

8a. Type of dwelling: 51

Detac ed, bungalow ..... -1
Semi-detac ed, corner ouse ..... -2
Terraced ouse ..... -3
Flat (upstairs or ground-floor), basement ..... -4
Part of business or commercial property (s op, office, works op, farm, market garden, etc.) ..... -5
Home for t e elderly, institution ..... -6
Ot er, viz.: ..... -7 52-53

[ ]

8b. How would you describe t e maintenance of t e *dwelling*:

KRT.04

E cellent ..... -1 54
Good ..... -2
Mediocre ..... -3
Poor ..... -4

8c. Location of main living room:

Basement ..... -1 55
Ground-floor ..... -2
Upper floor, viz.: ..... -3 56-57

8d. If code -4 was filled in for question 8a:

Number of floors of t e building occupied by t e dwelling: 58-59

8e. Satellite dis :

Visibly present ..... -1 60
Not observed ..... -2

Annex 7.1 Multinomial regression, logistic regression and discrete hazard rates

Multinomial regression

Multinomial logistic regression is a form of regression in which the dependent variable is nominal and distinguishes between different categories. The independent variables are generally nominal but can also be numerical variables. It is based on logits for each category of the dependent variable minus the 'reference group', so for a dependent with J categories, the logits are $\text{logit}(p_j)=\ln(p_j/(p_J)$. Of course it is immaterial which category is the J^{st} one. Logistic regression involves fitting the following equation, where x represents the independent variables:

$$\text{logit}(p_j) = a_j + \beta_{1j}x_1 + \beta_{2j}x_2 + \beta_{3j}x_3 + \beta_{3j}x_3 + \ldots$$

The output of the analysis comprises regression coefficients and standard errors of the independent variables for each of the categories of the dependent variable except the reference category. Put differently, logistic regression is the simple case of multinomial logistic regression (see below) in which there is only one set of parameters, namely for the non-reference category.

Table A6.1 gives an example of the results of a multinomial regression of the AVO-respondents. In this example the dependent variable has five classes and represents region. The fifth class (Amsterdam, Rotterdam and The Hague) is the reference category. The values under β for the intercept are the log odds-ratios of living in a particular region versus being a big-city resident (the reference category) if all independent variables are zero. β/S.E. is a z-statistic and has a standard normal distribution. Generally, the Wald statistic is presented, which is equal to $(\beta/\text{S.E.})^2$. The Wald statistic has a chi-square distribution with one degree of freedom. In the tables in this study the Wald statistic will be presented rather than the standard errors. The values under Significance show that most effects are strongly significant. The values under Exp(β) indicate the odds-ratios rather than log odds-ratios associated with each effect.

The chi-square value under model-fitting information is 1798 with 32 degrees of freedom (df) and is highly significant. This means that the null hypothesis that all effects of the independent variable are zero can be rejected. The likelihood ratio tests show that the null hypothesis that the effects on all four log odds-ratios of the dependent variable are simultaneously equal to zero can be rejected for the intercept and the independent variables. However, the loss of fit associated with 'high proportion of ethnic minorities' is much stronger than that of the other independent variables with 'no religious affiliation' as a second best. The contributions of the other independent variables are significant, but fairly small. The chi-square statistic in the likelihood ratio test is the difference in –2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0. The

reduced model for the intercept is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

The pseudo-R^2 measures indicate that the model performs fairly well. The Nagelkerke R^2 value will usually be the most relevant value to report. It corrects the Cox and Snell value so that it can theoretically achieve a value of 1. These pseudo-R^2 measures confound the goodness of fit and explanatory power of the model. Note that the pseudo-R^2 measures are not to be interpreted as the percentage of variance that is explained by the model, as they tend to underestimate this.

The β values in this example – which has no further theoretical significance – show that the major difference between big cities and the rest of the Netherlands is that inhabitants of big cities much more often live in neighbourhoods with a high rate of ethnic minorities (which of course does not come as a surprise). This is indicated by the size of the regression coefficients and of the Wald statistic. The beta for the north-eastern part of the Netherlands is most sizable, but also has a rather high standard error, so the Wald is smaller than for the other areas. Furthermore, a listed fixed phone number is more common outside big cities. Singles are less common outside big cities, young people more (at least among the AVO respondents). Outside big cities people also less often have no religious affiliation. The β values are significant at the 0.01 level in the mid-eastern part and the south of the Netherlands. People in big cities more often participate in cultural activities (again among the AVO-respondents). The pattern differs for participation in classical culture (concerts, theatre) and popular culture (pop concerts, cinema, dance).

Table 7.1 xample multinomial logistic regression: regions in the Netherlands (reference category: big cities) VO1999 (unweighted)

	west excl. big cities	north east	mid east	south	west excl. big cities	north east	mid east	south
	β				significance			
intercept	1.43	0.39	1.15	1.72	0.00	0.02	0.00	0.00
listed phone number	0.38	0.57	0.66	0.45	0.00	0.00	0.00	0.00
single	-0.36	-0.29	-0.27	-0.30	0.00	0.02	0.01	0.00
no religious affiliation	-0.12	-0.17	-0.63	-1.38	0.17	0.14	0.00	0.00
high rate ethnic minorities	-2.38	-3.96	-2.67	-2.72	0.00	0.00	0.00	0.00
cultural omnivore	-0.30	-0.41	-0.59	-0.63	0.01	0.00	0.00	0.00
classical only	-0.09	-0.06	-0.34	-0.51	0.51	0.73	0.02	0.00
popular only	-0.10	-0.30	-0.06	-0.17	0.47	0.09	0.68	0.25
	standard error				wald			
intercept	0.13	0.17	0.14	0.13	122	5	66	163
listed phone number	0.10	0.14	0.11	0.11	14	17	34	17
single	0.10	0.13	0.10	0.10	14	5	7	9
no religious affiliation	0.09	0.12	0.10	0.11	2	2	39	173
high rate ethnic minorities	0.10	0.26	0.12	0.12	573	240	478	514
cultural omnivore	0.11	0.15	0.12	0.12	7	8	23	27
classical only	0.13	0.16	0.14	0.14	0	0	6	14
popular only	0.14	0.18	0.15	0.15	1	3	0	1
	exp(β)							
listed phone number	1.46	1.77	1.94	1.57				
single	0.70	0.75	0.76	0.74				
no religious affiliation	0.88	0.84	0.53	0.25				
high rate ethnic minorities	0.09	0.02	0.07	0.07				
cultural omnivore	0.74	0.66	0.56	0.53				
classical only	0.92	0.95	0.71	0.60				
popular only	0.90	0.74	0.94	0.85				

Table 7.1 (cont.) xample multinomial logistic regression: regions in the Netherlands (reference category: big cities) V01999 (unweighted)

		%	likelihood ratio tests				
regions	Amsterdam, Rotterdam, The Hague (reference)	16.5	effect	-2 LL reduced	X^2	df	Sig.
	West (excl. big cities)	29.5	intercept	1088	0	0	
	north-east	10.4	listed phone	1124	37	4	0.00
	mid-east	20.3	single	1102	14	4	0.01
	south	23.4	no religion	1392	304	4	0.00
listed phone number	yes	78.8	ethnic minorities	2222	1134	4	0.00
	no (reference)	21.2	cultural participation	1148	60	12	
single	yes	27.0					
	no (reference)	73.0					
no religious affiliation	yes	34.9	**model fitting information**				
	no (reference)	65.1	model	-2 LL	X^2	df	Sig.
high rate ethnic minorities	yes	17.8	intercept only	2871			
	no (reference)	82.2	final	1088	1783	28	0.00
cultural participation	omnivore	30.6					
	classical only	22.4	**pseudo R^2**				
	popular only	16.7	Cox and Snell		0.253		
	no culture (reference)	30.4	Nagelkerke		0.265		
valid (N=6107)		100.0	McFadden		0.094		

Logistic regression

Logistic regression is generally used when the dependent variable can take on two values, for instance 'contact' and 'non-contact' or phone number available 'yes' or 'no'. It is the simple case of multinomial regression where one category is fitted (often the 'yes' category) and the other category is the reference value. Table A7.2 (taken from table 7.3) illustrates how logistic regression works. The substantive interpretation will be presented in chapter 6.

Table 7.2 Logistic regression availability listed phone number, VOA respondents

	%	%phone	β	exp β	standard error	Wald	sign. (df=1)
Amsterdam, Rotterdam, The Hague	16	66	–0.433	0.649	0.094	21.00	0.000
poor maintenance	20	65	–0.584	0.557	0.077	58.19	0.000
ethnic minorities 16+ %	18	65	–0.351	0.704	0.094	14.11	0.000
detached dwelling	14	93	0.918	2.504	0.141	42.30	0.000
ethnic minority group[a]	3	34	–1.331	0.264	0.185	51.53	0.000
family composition (df=2)						18.06	0.000
single	27	74	–0.160	0.853	0.089	3.21	0.073
not single, no children	34	84	0.213	1.237	0.085	6.22	0.013
not single, children	38	77					
higher education	18	83	0.417	1.518	0.097	18.50	0.000
sports activities (1 or more)	60	80	0.208	1.232	0.074	7.87	0.005
household owns pc	56	80	0.196	1.216	0.080	6.07	0.014
watches sport on tv	63	80	0.131	1.139	0.068	3.63	0.057
	mean	**mean phone**					
age	49.3	50.7	0.026	1.026	0.002	113.88	0.000
membership organization	1.42	1.48	0.044	1.045	0.024	3.40	0.065
constant			–0.107	0.899	0.155	0.47	0.491
total (n = 100%)	n = 6125	79					
model	X^2 (df=13)	634		Cox & Snell R^2		0.098	
–2 Log likelihood	5695			Nagelkerke R^2		0.153	

a Country of birth respondent, respondent's mother or respondent's father Morocco, Turkey, Surinam, Dutch Antilles or Aruba.

The dependent variable here is 'telephone number available' for respondent households. This holds for 79% of the households (p=0.79). Logistic regression is based on the logit(p) = ln(p/(1-p)). Logistic regression involves fitting the following equation:

$$\text{logit}(p) = a + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_3 x_3 + \ldots$$

This model is fitted using the maximum likelihood method, which maximizes the probability of obtaining the observed results given the fitted regression coefficient. The x are the independent variables. The β values are regression coefficients. If an x is numerical (e.g. age), the beta indicates that for each additional year the (logged) odds (of having a telephone number available rather than not) increase by .026 (=exp(0.026)), assuming that the household belongs to the reference group with regard to all other aspects. For the nominal variables the interpretation is as follows.

The odds of someone from an ethnic minority group having a listed phone number are 0.26 (exp(-1.331)), compared with those not from ethnic minority groups (and again assuming that the household belongs to the reference group in all other aspects). The fit of the model is indicated by –2 log-likelihoods, which compares the model with the L_0-model (constant only). The likelihood-ration chi-square is defined as $2(L_1 - L_0)$, where L_0 represents the log likelihood for the 'constant-only' model and L_1 is the log likelihood for the full model with constant and predictors. The Cox and Schnell and Nagelkerke R^2's indicate the goodness of fit. As in the case of multinomial regression they tend to underestimate the amount of variance explained.

Discrete hazard rates

The probability of contact being established at a particular call (see chapter 6), assuming noncontact prior to that call, can be modelled as a discrete hazard rate logistic model (see Campanelli, Sturgis and Purdon, 1997, p. 3-11). The hazard rate is defined as

$$\lambda(x|t) = P(T = t|T \geq t, x) = \frac{\exp(\beta_{0t} + x'\beta)}{1 + \exp(\beta_{0t} + x'\beta)}; t = 1,\ldots,q$$

where t is the call number and x are the predictors of contact at call τ. This model can be fitted in SPSS as a simple logistic regression by splitting up the original data file and setting up each call as a separate record. The dependent variable is contact (yes/no) and for one household there can be none, one or more noncontacts and only one contact (when analysing the likelihood of contacting previously noncontacted households).

Annex 9.1 Categorical Principal Components Analysis (CATPCA)

The Data Theory Scaling System Group (DTSS) of the Faculty of Social and Behavioral Sciences of Leiden University has developed a wide range of optimal scaling models and algorithms that are available in SPSS. In this study CATPCA is used, CATPCA is a categorical variant of Principal Components Analysis (PCA). In PCA weights are computed for each variable in a set in order to maximize the correlation of the original variables with the weighted total. Subsequently, component weights for further dimension independent of the former one can be computed. In CATPCA categorical variables are analysed in a similar way. Measurement restrictions per variable indicate whether these variables should be considered nominal, ordinal or numerical. These measurement restrictions determine the re-quantifications or transformations that are allowed for the categories of each variable. Multiple transformation of each variable are allowed which means that in each subsequent dimension a different transformation, within the measurement restrictions can be used, in order to maximize the item-total correlation. Slightly more complicated are single transformations which means that the transformation is identical in each dimension. In this cases dimensions are not computed sequentially, but a single solution is found in the pre-specified number of dimensions. In the figures presented in this study the transformations of the variables are given in a two-dimensional space. The category points of ordinal variables are linked, as are the category points of nominal variables that are ordered in real life but not in the analysis β (age).

The points in a plot of the results of CATPCA indicate categories of respondents. If two categories from different variables are close, this indicates that respondents that score in one category are likely to score in the other categories two, or that persons in both categories have similar scoring patterns on other variables. If two categories of the same variable are close, this indicates that persons in these categories have similar scoring patterns on other variables. Categories in the centre of the plot either do not fit well and have no clear-cut relationship with other variables, or they represent the vast majority of the individuals, so their centroid is the centroid of the plot.

Annex 10.1 Sections on response in the ESS specification for participating countries Round 1 (www.europeansocialsurvey.org Project Specifications)

5.6 Target response rates

Outcomes of all approaches to addresses, households and individuals in the sample will be defined and recorded according to a pre-specified set of categories that distinguish non-eligibility, non-contacts and refusals. The proportion of non-contacts should not exceed 3 per cent of all sampled units, and the minimum target response rate – after discounting ineligibles (and other 'deadwood', as defined by the CCT) – should be 70%. This figure is likely to be exceeded in certain countries and the ESS as a whole would be damaged if major national variations in response rates were to occur. Survey organisations should thus cost their surveys with this response rate in mind and consider what steps may be required to achieve it.

5.7 Response rate enhancement

Various specific steps designed to enhance response rates should also be allowed for. They include at least four personal visits by interviewers to each sampling unit before it is abandoned as non-productive, including at least one call in the evening and at least one at the weekend. These calls should be spread over at least two different weeks. Similarly, to allow difficult-to-contact people to be located, the fieldwork period should not be less than 30 days (see 5.13 and 5.15). All potential survey organisations must be invited to suggest a range of techniques that they believe would enhance the final response rate. Such techniques may include advance letters, toll-free telephone numbers for potential respondents to contact, extra training of interviewers in response-maximisation techniques and doorstep interactions, implementing refusal avoidance and conversion techniques, re-issuing of 'soft' refusals and 'indecisive' non-contacts, and many others. These suggestions will be forwarded to the CCT both for comment and for possible transmission to other nations for consideration locally.

5.8 Response rate calculation and documentation

Reporting of response outcomes will be carried out, calculated and keyed according to a pre-specified standard format, which will include at least the following mutually-exclusive categories:

A) *Total issued addresses* (or other sample units)
B) % not eligible, and why
C) *Total eligible sample* (A-B)
D) % no contact (after 4+ visits, or if fewer, why)
E) % personal refusal, and why (pre-specified categories)

G) % too ill or incapacitated
H) % household (or proxy) refusal, and why (pre-specified categories)
I) % achieved interview (partial & complete)

J) *Total percent response rate* (I/C)

5.9 *Field outcomes*

In addition to the recording of case outcomes, field outcomes for each call at each address (or other primary unit) will be documented and ultimately keyed from a standardized set of similar summary codes for each call. Interviewers will also be asked to record for each sample unit a number of observable area, dwelling and household characteristics for each case in the issued sample (to be specified in detail later).

Annex 10.2 Response enhancement in the European Social Survey Round 1 2002/2003

country	incentive	refusal conversion
Austria	no	second contact
Belgium	no	Based on contact form data, selection of respondents likely to cooperate. For example those which refused because lack of time, too busy, those estimated by the interviewer as likely to cooperate in future rounds. Re-issue of soft refusal (new interviewer).
Czech Republic	no	no
Denmark	no	
Finland	no	The refusers were sent a new letter and/or contacted by the same or a different interviewer. Strategies for each refusal were discussed in an interviewer meeting to minimize respondent refusal.
France	no	no
Germany	gifts: Electronic calculator with ess-Logo, raffle of 5 weekend trips to European capitals	... we tried to "convert" the refusals into realized interviews and to get in touch with the non-contacts. A CATI-questionnaire was developed for "refusal conversion". Out of 482 refusals, 95 interviews could be realized. The non-contacts were also tried to be contacted by telephone. The central aim was to find out about the reachability of these persons (for personal contacts with the interviewers).
Greece	A bag with a set of pens & mechanical pencils, 2 discount coupons for orange juice, 1 discount coupon for milk and 2 discount coupons for chocolate milk.	An advance letter at the household's main entrance during recording of all households in sampling area. A second letter was handed to respondent in case of refusal or placed under door upon prolonged absence. All refusals were reissued and senior interviewers were used. A phone number was stated at respondent letters inducing call to National Coordinator. A researcher at EKKE premises was handling those phone calls to reduce reluctance or cater queries.
Hungary	no	Interviewers attempted to describe the content of the survey, to emphasize its international character. They tried to make the respondent understand that participation of all randomly selected person is very important.
Ireland	no	no
Israel	lottery tickets (IS 20)	First trial by interviewer. If unsuccessful the interviewer gave details to their supervisor who contacted respondent by phone. If successful, a time was set for interviewer to return. In some cases the case was referred to the central office for further attempts.
Italy	gift	green number for any kind of information
Luxemburg	money EUR 20	no

Annex 10.2 Response enhancement in the European Social Survey Round 1 2002/2003

country	incentive	refusal conversion
Netherlands	From week 37 on, i.e. the start of the fieldwork, respondents were encouraged to participate with one choice out of three presents (worth EUR 2.50). From week 49 on, respondents were encouraged with financial donations (worth EUR 5) plus a quiz with prizes to be won (1*EUR 250, 3*EUR 100).	Interviewers were provided during interviewer's instructions with arguments to persuade respondents to participate. They carried a brochure with background information on ess and they could refer to an ess-website with additional information. Halfway through the interview period, a second advance letter was sent and previously refusing respondents were phoned to ask which letter they liked most and were then again asked to participate.
Norway	all respondents received a lottery ticket in the advance letter	The interviewers were instructed and trained to listen to the type of concerns the respondent communicated, then identify a way to address them and deliver an answer quickly in the respondent's native language. E.g.: Respondent: Don't you know there's a soccer game on right now? Interviewer: Then this is probably a bad time for you, would it be OK if I call you back when the game is finished?
Poland	no	In case of 'soft refusal' (if the respondent hesitated or refusal was declared by other person, etc.) the interviewer visited respondent once again.
Portugal	no	new contact by letter and / or phone
Slovenia	no	additional follow up letter, re-issuing with an elite group of interviewers.
Spain	one T-shirt and cap per respondent.	no
Sweden	lottery ticket to all who received the advance letter, whether they were interviewed or not	repeated contacts
Switzerland	ten special edition priority stamps (heart-shaped)	Re-contact by the best interviewers with detailed information about the reasons for the previous refusal. Re-contact by telephone from the call centre.
UK	£5 gift voucher - mentioned in advance letter and handed out by interviewer at end of interview	Interviewers told to withdraw if a refusal seemed imminent and to return another day. Most refusals were reissued to another interviewer to try again.

Source: NSD (2004)

References

Aarts, Kees, Bas Denters and Ingrid Smeets (1998) *Built on Quicksand? Survey Response Bias and Political Interest.* Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago Ill., April 1998.

Aarts, Kees & Henk van der Kolk (1999) *Tegen de stroom in? Non-respons en Hawthorne effecten in het Nationaal Kiezersonderzoek 1998*. Paper presented at the NVMC-voorjaarsconferentie 'De waan van de Dag', Utrecht, April 1999.

Atrostic, B.K., Nancy Bates, Geraldine Burt and Adriana Silberstein (2001) Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends and New Insights. *Journal of Official Statistics*, Vol. 17, N° 2, pp. 209-226.

Allison, Paul D. (2002) *Missing Data*. California: Sage Publications.

American Statistical Association (2000) *What is a Survey?* ASA Series, Washington, DC: ASA. www.amstat.org/sections/srms (last modification 21 May 2000)

American Association for Public Opinion Research (2000) *Standards and Best Practices*. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Ann Arbor, Michigan: AAPOR. www.aapor.org/default.asp?page=survey_methods/standards_and_best_practices/standard_definitions

Andeweg, Rudy B., and Joop van Holsteyn (1996) A Hidden Confidence Gap? The Question of Nonresponse Bias in Measuring Political Interest. *The Netherlands' Journal of Social Sciences*, Vol. 32, N° 2, December 1996, pp. 127-142.

Babbie, Earl (1992) *The Practice of Social Research* (6th ed.) Belmont, CA: Wadsworth.

Bates, N. and K. Creighton (2000) *The Last Five Percent: What Can We Learn from Difficult Interviews?* Proceedings of the Annual Meetings of the American Statistical Association, August 2000. www.fcsm.gov/committees/ihsng/asa2000proceedings.pdf

Becker, H.A. (1972) Observation by Informants in institutional research. *Quality and Quantity*, Vol. VI, N° 1, pp. 157-169.

Bethlehem, Jelke (2000) De Klassieke Steekproeftheorie – Een Overzicht. *Kwantitatieve Methoden*, jr. 21, nr. 65, pp. 69-84.

Bethlehem, Jelke G. (2002) Weighting Nonresponse Adjustments Based on Auxiliary Information. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 265-287.

Bethlehem, J.G., & H.M.P. Kersten (1986) *Werken met non-respons*. Statische Onderzoekingen M30, CBS-publikaties, 's-Gravenhage: Staatsuitgeverij.

Bethlehem, Jelke, and Barry Schouten (2003) *Nonresponse Analysis of the Integrated Survey on Living Conditions (POLS)*. Paper presented at the 14th International Workshop on Household Survey Nonresponse, Leuven, September 2003.

Biemer, Paul P. (2001) Comment (on: Can a Statistician Deliver, JOS, same issue) *Journal of Official Statistics* Vol. 17, N° 1, pp. 25-32.

Biemer, Paul P., and Lars E. Lyberg (2003) *Introduction to Survey Quality*. Hoboken, New Jersey: John Wiley and Sons, Inc.

Biemer, Paul P., Henry Woltmann, David Raglin and Joan Hill (2001) Enumeration Accuracy in a Population Census: An Evaluation Using Latent Class Analysis. *Journal of Official Statistics*, Vol. 17, N° 1, pp. 129-148.

Billiet, Jaak, and Michel Philippens (2004) *Data Quality Assessment in ESS Round 1: Between Wishes and Reality*. Paper presented at the RC33 Sixth International Conference on Social Science Methodology, Amsterdam, August 2004. ess.nsd.uib.no/2003_documentation.jsp (Data Quality)

Blohm, Michael, und Claudia Diehl (2001) Wenn Migranten Migranten befragen. Zum Teilnahmeverhalten von Einwanderern bei Bevölkerungsbefragungen. *Zeitschrift für Soziologie*. Jg. 30, Heft 3, Juni 2001, pp. 223-242.

Bogen, Karen (1996) *The Effect of Questionnaire Length on Response Rates - A Review of the Literature*. Proceedings of the Survey Research Methods Section, American Statistical Association. www.census.gov/srd/papers/pdf/kb9601.pdf

Bowers, Diane (2004) The US Do-Not-Call registry: A victory for the US research industry. *Research World*, June 2004, p. 17.

Borg, Ingwer (2000) Früh- versus Spätantworter. *ZUMA-Nachrichten* 47, Jg. 24, November 2000, pp. 7-19.

Bradburn, Norman N. (1992) Presidential Address: A Response to the Nonresponse Problem. *Public Opinion Quarterly*, Vol. 56, pp. 391-397.

Brehm, John (1993) *The Phantom Respondents: Opinion surveys and political representation*. Ann Arbor: University of Michigan Press.

Bronner A.E. (1988) Surveying Ethnic Minorities. In: W.E. Saris and I.N. Gallhofer (eds.) *Sociometric Research*, Vol. 1. London: MacMillan, pp. 36-47.

Bronner, Fred, Costa Tchaoussoglou and Raymond Ross (2003) *The Virtual Interviewer*. Paper presented at the 11th Worldwide Readership Symposium, Basel, October 2003.

Brunner, G. Allen, and Stephen J. Carroll, Jr. (1967) The Effect of Prior Telephone Appointments on Completion Rates and Response Content. *Public Opinion Quarterly*, Vol. 31, pp. 652-654.

Campanelli, Pamela, Patrick Sturgis and Susan Purdon (1997) *Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates.* London: The Survey Methods Centre at SCPR.
www.soc.surrey.ac.uk/~scs1ps/papers/can you hear me knocking.pdf

Carton, Ann (1999) *Selectie, training en evaluatie van interviewers binnen een interviewernetwerk*. Leuven: Garant.

Carton, A. (2001) Total Quality Management bij training en supervisie van interviewers. In: A.E. Bronner et al. (eds.) *Ontwikkelingen in het marktonderzoek*. Jaarboek 2001 Marktonderzoek Associatie. Haarlem: De Vrieseborch, pp. 105-121.

Centraal Bureau voor de Statistiek (1987) *De leefsituatie van Turken en Marokkanen in Nederland 1984*. 's-Gravenhage: Staatsuitgeverij.

Centraal Bureau voor de Statistiek (1991) *De gezondheidsenquête Turkse ingezetenen in Nederland 1989/1990*. Voorburg/Heerlen: CBS.

Centraal Bureau voor de Statistiek (2000) *Statistisch Jaarboek 2000*. Voorburg/Heerlen: CBS.

Centraal Bureau voor de Statistiek (2001) Bevolking in institutionele huishoudens, 1 januari 2000. *Maandstatistiek van de bevolking*, 5, p. 39.

Clarridge, Brian, and Michael P. Massagli (1995) Consequences of the Hard Pursuit of Non-respondents in Two Physician Studies. American Statistical Association, *Proceedings of the International Conference on Survey Measurement and Process Quality, Contributed Papers*. Bristol, 1995, Alexandria: ASA, pp. 103-108.

Cohen, G., and J.C. Duffy (2002) Are Nonrespondents to Health Surveys Less Healthy Than Respondents? *Journal of Official Statistics*, Vol. 18, N° 1, pp. 13-23.

Converse, Jean M. (1987) *Survey Research in the United States. Roots & Emergence 1890 – 1960*. Berkeley, University of California Press.

Converse, Jean M., and Howard Schumann (1974) *Conversations at Random. Survey Research as Interviewers See It.* New York: John Wiley and Sons, Inc.
Couper, Mick P. (1997) Survey Introductions and Data Quality. *Public Opinion Quarterly*, Vol. 61, pp. 317-338.
Couper, Mick P. (2000) Web Surveys. A Review of Issues and Approaches. *Public Opinion Quarterly*, Vol. 64, pp. 464-494.
Couper, Mick P. (2001) The Promises and Perils of Web Surveys. In: Andrew Westlake, Wendy Sykes, Tony Manners and Malcom Riggs (eds.) *The Challenge of the Internet*. Proceedings of the ASC International Conference on Survey Research Methods. Chesham, UK, May 2001.
Couper, Mick P. (2002) *New Technologies and Survey Data Collection: Challenges and Opportunities*. Invited paper presented at the International Conference on Improving Surveys, Copenhagen, August 2002.
www.icis.dk/ICIS_papers/Keynote1_0_3.pdf
Couper, Mick P., and Edith D. de Leeuw (2003) Nonresponse in Cross-Cultural and Cross-National Surveys. In: Janet A. Harkness, Fons J.R. van de Vijver and Peter Ph. Mohler (eds.) *Cross-Cultural Survey Methods*. New York: Wiley, pp. 157-177.
Couper, Mick P., and Robert M. Groves (1996) Social environmental impacts on survey cooperation. *Quality and Quantity*, Vol. 30, pp. 173-188.
Couper, M.P., E. Ryu, E. and R.W. Marans (2002) *Incentives: Cash Versus In-Kind, Face-to-Face Versus Mail, Response Rate Versus Nonresponse Bias*. Paper presented at the13th International Workshop on Household Survey Nonresponse, Copenhagen, August 2002.
Curtin, Richard, Stanley Presser and Eleanor Singer (2000) The Effects of response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, Vol. 64, pp. 413-428.
Däubler, Thomas (2002) Nonresponseanalysen der Stichprobe F. Materialen 15. Berlin: Deutsches Institut für Wirtschaftsforschung. www.diw.de/deutsch/publikationen/materialien/docs/papers/diw_rn02-05-15.pdf.
Davern, Michael, Todd H. Rockwood, Randy Sherrod and Stephen Campbell (2003) Prepaid Monetary Incentives and Data Quality in Face-to-face Interviews. Data from the 1996 Survey of Income and Program Participation Incentive Experiment. *Public Opinion Quarterly*, Vol. 67, pp. 139-147.
De Kruijk, M., & E. Hermans (1998) Effect van non-response op de representativiteit. In: A.E. Bronner et al. (eds.) *Recente ontwikkelingen in het marktonderzoek*. Jaarboek van de Nederlandse Vereniging voor Marktonderzoek en Informatiemanagement. Haarlem: De Vrieseborch, pp. 55-69.
De Haan, Jos (2001) Uitgaan, het bezoek aan horeca, cultuur en sportwedstrijden. In: Koen Breedveld & Andries van den Broek (eds.) *Trends in de tijd*. Den Haag: Sociaal en Cultureel Planbureau.
www.scp.nl/boeken/titels/2001-5/nl/metainfo.htm
De Haan, J.P., & W. Knulst (2000) *Het bereik van de kunsten*. Den Haag: Sociaal en Cultureel Planbureau.
www.scp.nl/boeken/cahiers/cah166/nl/metainfo.htm
De Heer, Wim (1999) International Response Trends, Results of an International Survey. *Journal of Official Statistics*, Vol. 15, N° 2, pp. 129-142.
De Heer, Wim (2000) Survey Practices in European Countries. In: Siobhán Carey (ed.) *Measuring Adult Literacy. The International Adult Literacy Survey (IALS) in the European context*. London: Office for National Statistics, pp. 43-64.

De Heer, Wim, & Edith D. de Leeuw (1999) Enquêteren bij personen en huishoudens, een bewogen geschiedenis. In: B. Erwich & J.G.S.J. van Maarseveen, *Een eeuw statistieken. Historisch-methodologische schetsen van de Nederlandse officiële statistieken in de twintigste eeuw*. Voorburg, CBS, pp. 47-70.

De Heer, Wim, Edith D. de Leeuw and Johannes van der Zouwen (1999) Methodological Issues in Survey Research: A Historical Review. *Bulletin de Méthodologie Sociologique*, October 1999, n. 64, pp. 25-48.

Dekker, P. (ed.) (2002) *Niet-stemmers: Een onderzoek naar achtergronden en motieven in enquêtes, interviews en focusgroepen*. Den Haag: Sociaal en Cultureel Planbureau. www.scp.nl/boeken/titels/2002-6/nl/metainfo.htm

De Leeuw, Edith D. (1992) *Data Quality in Mail, Telephone and Face to Face Surveys*. Amsterdam: TT-Publikaties.

De Leeuw, E.D. (1997) Nonrespons: Wat weten we en wat kunnen we eraan doen. In: A.E. Bronner et al. (eds.) *Recente ontwikkelingen in het marktonderzoek*. Jaarboek van de Nederlandse Vereniging voor Marktonderzoek en Informatiemanagement. Haarlem: De Vrieseborch, pp. 39-59.

De Leeuw, Edith D. (1999) How do Successful and Less Successful Interviewers Differ in Tactics for Combating Survey Nonresponse? *Bulletin de Méthodologie Sociologique*, April 1999, n. 62, pp. 29-42.

De Leeuw, Edith (2001) I am not selling anything: Experiments in telephone introductions. *Kwantitatieve Methoden*, jaargang 22, nr. 68, pp. 41-48.

De Leeuw, Edith, and Wim de Heer (2002) Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 41-54.

De Leeuw, Edith D., and Joop J. Hox (1998) Nonrespons in surveys: een overzicht. *Kwantitatieve Methoden*, jaargang 19, nr. 57, pp. 31-53.

De Leeuw, Edith, Joop Hox, Ger Snijkers and Wim de Heer (1998) Interviewer Opinions, Attitudes and Strategies Regarding Survey Participation and Their Effect on Response. In: A. Koch & R. Porst (eds.) *Nonresponse in Survey Research*. ZUMA-Nachrichten Spezial Band 4. Mannheim: ZUMA.

De Leeuw, Edith, Jim Lepkowski and Sun-Woong Kim (2002) *Have Telephone Surveys a Future in the 21st century?* Paper presented at the International Conference on Improving Surveys. Copenhagen, August 2002.

de Vaus, David (2002) Editor's Introduction : Social Surveys – An Overview. In: David de Vaus (ed.) *Social Surveys*. Sage Benchmarks in Social Research Methods (4 Vols.) London: Sage Publications, pp. IX-XLIV. users.bigpond.net.au/david-devaus/My-Books/4Volumes/surveys%204%20volumes_Introduction.pdf_2.pdf

Deming, W. (1960) *Sample design in business research*. New York: Wiley.

Deville, Jean-Claude (1991) A Theory of Quota Surveys. *Survey Methodology*. Vol. 17, N° 2. pp. 163-181.

Deutskens, E.C., P.M. Willems, S.S.M. van den Berg, P. Oosterveld, J.C. de Ruyter & M.G.M. Wetzels (2003) Conversie van gestandaardiseerd vragenlijstonderzoek naar internet: kritische succesfactoren. In: A.E. Bronner et al. (eds.) *Ontwikkelingen in het marktonderzoek*. Jaarboek 2003 Marktonderzoek Associatie. Haarlem: De Vrieseborch, pp. 41-58.

Diffendal, G. (2001) *The Hard to Interview in the American Community Survey*. Proceedings of the Annual Meetings of the American Statistical Association, August 2001. www.fcsm.gov/committees/ihsng/diffendalasa2000.pdf

Dijkstra, Wil, and Johannes H. Smit (2002) Persuading Reluctant Recipients in Telephone Surveys. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 121-134.

Dillman, Don A. (1999) *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.

Dillman, Don A. and Dennis K. Bowker (2001) The Web Questionnaire Challenge to Survey Methodologists. In: Ulf-Dietrich Reips & Michael Bosnjak (eds.) *Dimensions of Internet Science*. Pabst Science Publishers www.pabstpublishers.com ; survey.sesrc.wsu.edu/dillman/papers.htm.

Dillman, Don A., John L. Eltinge, Robert M. Groves and Roderick J.A. Little (2002) Survey Nonresponse in Design, Data Collection, and Analysis. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 3-26.

Dillman, Don A., Jean Gorton Gallegos and James H. Frey (1976) Reducing Refusal Rates for Telephone Interviews. *Public Opinion Quarterly*, Vol. 40, N° 1, pp. 66-78.

Doherty, M. (1994) Probability versus Non-Probability Sampling in Sample Surveys. *The New Zealand Statistics Review*. March 1994 issue, pp. 21-28. www.sch.abs.gov.au/SCH%5CA1610103.NSF/0/1BDA1D5801E12326CA2569A600002192?OpenDocument

Drago, Robert, Robert Caplan, David Castanza, Tanya Brubaker, Darnell Cloud, Susan Donohue, Naomi Harris and Tammy Riggs (1999) Time for Surveys: Do Busy People Complete Time Diaries? *Loisirs & Société*, Vol. 21, n° 2, Fall 1998, pp. 555-563.

Duhart, D., N. Bates, B. Williams, G. Diffendal and P. Chiu (2001) *Are Late/Difficult Cases in Demographic Survey Interviews Worth the Effort? A Review of Several Federal Surveys*. Proceedings of the Federal Committee on Statistical Methodology Research Conference, November 2001. www.fcsm.gov/committees/ihsng/Duhart.pdf

Ellemers, J.E. (2004) Wat we nog niet weten. Tekortkomingen en blinde vlekken bij de analyse van de Nederlandse maatschappij. *Sociologische Gids* 2004/4 (jg 51) pp. 482-487.

Elliot, Dave (1991) *Weighting for non-response. A survey researcher's guide*. London: Office of Population Censuses and Surveys.

Esomar (2003a) Research in the Netherlands – The Strength is in online research. ESOMAR *Research World*, Vol. 11 N° 1, January 2003, pp. 4-6.

Esomar (2003b) Research Issues – The 'Do not call register' in the US. ESOMAR *Research World*, Vol. 11 N° 11, December 2003, pp. 22.

ESS (2004) *Field Procedures in the European Social Survey: Enhancing Response Rates*. ESS-documents (26/07/04). http://naticento2.uuhost.uk.uu.net/fieldwork/response_enhancement_round_2.doc.

Faas, Thorsten (2003) Offline rekrutierte Access Panels: Königsweg der Online-Forschung? ZUMA-*Nachrichten* 53, Jg. 27, November 2003, pp. 58-76 www.gesis.org/Publikationen/Zeitschriften/ZUMA_Nachrichten/documents/pdfs/53/08_Faas.pdf.

Federal Committee on Statistical Methodology (2001) *Measuring and Reporting Sources of Errors in Surveys*. Statistical Policy Working Paper 31. www.fcsm.gov/01papers/SPWP31_final.pdf.

Fellegi, Ivan P. (2001) Comment (on: Can a Statistician Deliver, JOS, same issue) *Journal of Official Statistics* Vol. 17, N° 1, pp. 43-50.

Fowler, Floyd J. Jr. (2002) *Survey Research Methods*, 3rd Edition. Thousand Oaks, California: Sage Publications.

Fraboni, Romina, Alessandro Rosina, Sante Orsini and Barbara Baldazzi (2002) *A Multilevel Approach to the Analysis of Refusal Risk in the Italian Multipurpose Survey*. Paper presented at the International Conference on Improving Surveys, Copenhagen, August 2002. www.icis.dk/ICIS_papers/D1_3_4.pdf

GfK (1999) *GfK MiniCensus*. www.gfk.nl/data/consscop/text/minicens.htm.
GfK (2000) *Onderzoeksverantwoording Aanvullend Voorzieningengebruikonderzoek 1999*. Dongen: GfK www.scp.nl/miss/avodoc/avo99ov.pdf
GfK (2002) *Onderzoeksverantwoording Non-responsonderzoek*. Dongen: GfK
GfK (2003) *Onderzoeksverantwoording Aanvullend Voorzieningengebruikonderzoek 2003*. Dongen: GfK.
Gorden, Raymond L. (1980) *Interviewing: Strategy, Techniques and Tactics* (3rd edition). Homewood, Ill: Dorsey Press.
Goyder, John (1986) Surveys on Surveys: Limitations and Potentialities. *Public Opinion Quarterly*, Vol. 50, pp. 27-41.
Goyder, John (1987) *The Silent Minority. Nonrespondents on Sample Surveys*. Cambridge: Polity Press.
Goyder, John, Jean Lock and Trish McNair (1992) Urbanization effects on survey nonresponse: a test within and across cities. *Quality & Quantity*, Vol. 26, pp. 39-8.
Goyder, John, Keith Warriner and Susan Miller (2002) Evaluating Socio-economic Status (SES) Bias in Survey Nonresponse. *Journal of Official Statistics*, Vol. 18, Nº 1, pp. 1-11.
Groves, Robert M. (1989) *Survey Errors and Survey Costs*. New York: John Wiley & Sons, Inc.
Groves, Robert M., Paul Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls, II, and Joseph Waksberg (2001) *Telephone Survey Methodology*. New York: Wiley.
Groves, Robert M., Robert B. Cialdini and Mick P. Couper (1992) Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly*, Vol. 56, pp. 475-495.
Groves, Robert M., and Mick P. Couper (1998) *Nonresponse in Household Interview Surveys*. New York: Wiley.
Groves, Robert M., and Katherine A. McGonagle (2001) A Theory-Guided Interview Training Protocol Regarding Survey Participation. *Journal of Official Statistics*, Vol. 17, Nº 2, pp. 249-266.
Groves, Robert M., Stanley Presser and Sarah Dipko (2004) The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly*, Vol. 68, pp. 2-31.
Groves, Robert M., Eleanor Singer and Amy Corning (2002) Leverage-Saliency Theory of Survey Participation. Description and an Illustration. *Public Opinion Quarterly*, Vol. 64, pp. 299-308.
Groves, R.M. and D. Wissoker (1999) *Early Nonresponse Studies of the 1997 National Survey of America's Families*. Washington, DC: Urban Institute, report Nº 7. www.urban.org/UploadedPDF/Methodology_7.pdf
Groves, Robert M., and Sonja I. Ziniel (2003) *Using Interviewer Observations as Predictors of Contactability in US Face to Face Surveys*. Paper presented at the 14th International Workshop on Household Survey Nonresponse, Leuven, September 2003.
Guadagnoli, Edward, and Susan Cunningham (1989) The Effects of nonresponse and late response on a survey of physician attitudes. *Evaluation & the Health Professions*, Vol. 12, Nº 3, September 1989, pp. 318-328.
Hansen, M., and W. Hurvitz (1946) The problem of nonresponse in sample surveys. *Journal of the American Statistical Association*, Vol. 41, pp. 517-529.
Heeg, Robert (2004) The search for talent: 'A career in market research is a hard sell'. *Research World*, September 2004, p. 10-12.
Hippler, Hans-J., and Gabriele Hippler (1986) Reducing Refusal Rates in the Case of Threatening Questions: The 'Door-in-the-Face' Technique." *Journal of Official Statistics*, Vol. 2, Nº 1, pp. 25-33.
Holbrook, A. L., M. Green and J.A. Krosnick (2003) Telephone vs. face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, Vol. 67, pp. 79-125.

Houtkoop-Steenstra, Hanneke, and Huub van den Bergh (2000) Effects of Introductions in Large-Scale Telephone Survey Interviews. *Sociological Methods & Research*, Vol. 28, N° 3, February 2000, pp. 281-300.

Hox, Joop. J., and Edith D. de Leeuw (1994) A comparison of nonresponse in mail, telephone, and face-to-face surveys: Applying multilevel modeling to meta-analysis. *Quality and Quantity*, Vol. 28, pp. 329-344.

Hox, Joop, Edith de Leeuw and Harrie Vorst (1995) Survey Participation as Reasoned Action; a Behavioral Paradigm for Survey Nonresponse? *Bulletin de Méthodologie Sociologique*, N° 48, September 1995, pp. 52-67.

Hox, Joop, and Edith de Leeuw (2002) The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse: An International Comparison. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 103-120.

INFAS (2003) *Methodenbericht European Social Survey*, Deutsche Teilstudie. Bonn, July 2003.

Inter/view-groep (1978) *Rondje Non-response van de Inter/view-groep*. Amsterdam: Inter/view B.V.

Japec, Lilli, and Peter Lundqvist (1999) *Interviewer Strategies and Attitudes.* Paper presented at the International Conference on Survey Nonresponse. Portland, Oregon, October 1999.

Jansma, F., H. van Goor & R. Veenstra (2003) Verschillen in stadswijken in onderdekking en nonrespons? Een verkennend onderzoek naar selectieve uitval in een telefonische enquête. In: A.E. Bronner et al. (eds.) *Ontwikkelingen in het marktonderzoek*. Jaarboek 2003 Marktonderzoek Associatie. Haarlem: De Vrieseborch, pp. 41-58.

Jowell, Roger (1998) How Comparative is Comparative Research? *American Behavioral Scientist*, Vol. 42, N° 2, October 1998, 168-177.

Kalfs, N. (2001) Van oude en nieuwe dataverzamelingstechnieken: onderzoek naar verplaatsingsgedrag. In: A.E. Bronner et al. (eds.) *Ontwikkelingen in het marktonderzoek*. Jaarboek 2001 Marktonderzoek Associatie. Haarlem: De Vrieseborch, pp. 89-104.

Kalfs, N., & E. Kool (1994) *Ervaringen met non-respons*. Amsterdam: NIMMO.

Kalton, Graham (1987) *Introduction to Survey Sampling*. California: Sage Publications.

Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves and Stanley Presser (2000) Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly*, Vol. 64, pp. 125-148.

Kennickell, Arthur B. (2000) *What Do the "Late" Cases Tell Us? Evidence from the 1998 Survey of Consumer Finances*. Paper presented at the International Conference on Survey Nonresponse. Portland, Oregon, October 1999 (revised version) www.federalreserve.gov/Pubs/OSS/oss2/papers/icsn99.9.pdf

Kish, Leslie (1994) *Survey Sampling*. New York: Wiley and Sons, Inc.

Kish, Leslie (1998) *Quota Sampling: Old Plus New Thought*. Working paper, University of Michigan.

Kish, Leslie (2003) The Hundred Years' War of Survey Sampling (Reprinted from Statistics in Transition, 1995). In: Graham Kalton and Steven Heeringa (eds.) *Leslie Kish, Selected Papers*. New York: Wiley, pp. 5-19.

Knulst, Wim, and Andries van den Broek (1999) Do Time-use Surveys Succeed in Measuring "Busyness"? Some Observations of the Dutch Case. *Loisirs & Société*, Vol. 21, n° 2, Fall 1998, pp. 563-572.

Koch, Achim (1993) Sozialer Wandel als Artefakt unterschiedlicher Ausschöpfung? Zum Einfluß von Veränderungen der Ausschöpfungsquote auf die Zeitreihen des ALLBUS. ZUMA-*Nachrichten*, 33, Jg. 17, November 1993, pp. 43-113.

Krosnick, Jon A. (1999) Survey Research. *Annual Review of Psychology*, 50, pp. 537-567.

Kruskal, William and Frederick Mosteller (1979a) Representative Sampling, I: Non-scientific Literature. *International Statistical Review* 47, pp. 13-24.

Kruskal, William and Frederick Mosteller (1979b) Representative Sampling, II: Scientific Literature, Excluding Statistics. *International Statistical Review* 47, pp. 111-127.
Kruskal, William and Frederick Mosteller (1979b) Representative Sampling, III: the Current Statistical Literature. *International Statistical Review* 47, blz. 245-265.
Kuusela, Vesa, and Matti Simpanen (2002) *Effects of Mobile Phones on Telephone Survey, Practice and Results*. Paper presented at the International Conference on Improving Surveys. Copenhagen, August 2002. www.icis.dk/ICIS_papers/A_2_3.pdf
Lamas, Carlos (2003) *The Respondent: an Endangered Species*. Paper presented at the CESP symposium <Qui répond aux enquêtes> Paris, November 2003. www.cesp.org
Laurie, Heather, Rachel Smith and Lynne Scott (1999) Strategies for reducing nonresponse in a longitudinal survey. *Journal of Official Statistics*, Vol. 15, N° 2, pp. 269-282.
Lemay, Michael and Claire Durand (2002) The Effect of Interviewer Attitude on Survey Cooperation. *Bulletin de Méthodologie Sociologique*, N° 76, October 2002, pp. 27-44.
Lievesley, Denise (1983) *Reducing Unit Non-response in Interview Surveys*. Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 295-299.
www.amstat.org/sections/srms/proceedings/papers/1983_056.pdf
Lievesley, D. (1986) *Unit Non-response in Interview Surveys*. Joint Centre for Survey Methods unpublished paper. London: SCPR.
Lin, I-Fen, and Nora Cate Schaeffer (1995) Using Survey Participants to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly*, Vol. 59, pp. 236-258.
Little, Roderick J.A., and Donald B. Rubin (1987) *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.
Loosveldt, Geert, and Ann Carton (2002) Utilitarian individualism and panel nonresponse. *International Journal of Public Opinion Research*, Vol. 14, N° 4, pp. 428-438.
Loosveldt, Geert, Michel Philippens, Ineke Stoop and Jaak Billiet (2003) *Refusal conversion procedures in the European Social Survey*. Paper presented at the 14th International Workshop on Household Survey Nonresponse, Leuven, September 2003. ess.nsd.uib.no/2003_documentation.jsp (Data Quality)
Loosveldt, Geert, and Vicky Storms (2001) *The relevance of respondents' attitudes towards surveys*. Paper presented at the 12th International Workshop on Household Survey Nonresponse, Oslo, September 2001.
Loosveldt, Geert, & Vicky Storms (2003) Peilen in Vlaanderen. De houding van de Vlaming t.a.v. surveyonderzoek. In: *Vlaanderen gepeild!*. Studie gehouden te Brussel op 6 mei 2003. Brussel: Ministerie van de Vlaamse Gemeenschap, pp. 347-370. aps.vlaanderen.be/statistiek/publicaties/pdf/survey/gepeild2003/survey.pdf
Louwen, Frans (1992) Bereidwillig maar nier bereikbaar of bereikbaar maar niet bereidwillig. *Onderzoek*, juli 1992, pp. 5-9.
Lyberg, Lars (ed.) (2001) 'Can a Statistician Deliver' and Comments, *Journal of Official Statistics*, Vol. 17, N° 1, pp. 1-127.
Lyberg, Lars, et al. (2001) *Summary Report from the Leadership Group* (LEG) on Quality, SPC.
Lynn, P. (1996) Weighting for survey non-response. In: R. Banks et al. (ed.s), *Survey and Statistical Computing 1996*, pp. 205-214. Association for Survey Computing. www.iser.essex.ac.uk/home/plynn/downloads/Lynn%201996%20Weighting.pdf
Lynn, Peter (2001) The Impact of Incentives on Response Rates to Personal Interview Surveys: Role and Perceptions of Interviewers. Research Note. *International Journal of Public Opinion Research*, Vol. 13 N° 3, pp. 326-336. ijpor.oupjournals.org/cgi/reprint/13/3/326.pdf.
Lynn, Peter (2002) PEDAKSI: Methodology for Collecting Data about Survey Non-respondents. *Working papers of the Institute for Social and Economic Research*, paper 2002-05. Colchester: University of Essex.
www.iser.essex.ac.uk/pubs/workpaps/pdf/2002-05.pdf.

Lynn, Peter (2003) PEDAKSI: Methodology for Collecting Data about Survey Non-Respondents. *Quality and Quantity*, Vol. 37, pp. 239-261.

Lynn, Peter (2004) The Use of Substitution in Surveys. *The Survey Statistician*, N° 49, January 2004, pp. 14-16.

Lynn, Peter, Roeland Beerten, Johanna Laiho and Jean Martin (2001) Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys. *Working Papers of the Institute for Social and Economic Research*, paper 2001-23. Colchester: University of Essex. www.iser.essex.ac.uk/pubs/workpaps/pdf/2001-23.pdf.

Lynn, Peter, Roeland Beerten, Johanna Laiho and Jean Martin (2002) Towards Standardisation of Survey Outcome Categories and Response Rate Calculations, *Research in Official Statistics*, 2002, 1, pp. 63-86.
europa.eu.int/comm/eurostat/Public/datashop/print-product/EN?catalogue=Eurostat&product=KS-CS-02-001-__-N-EN&mode=download

Lynn, Peter, and Paul Clarke (2001) Separating Refusal Bias and Non-contact Bias: Evidence from UK National Surveys. *Working Papers of the Institute for Social and Economic Research*, paper 2001-24. Colchester: University of Essex. www.iser.essex.ac.uk/pubs/workpaps/pdf/2001-24.pdf.

Lynn, Peter, Paul Clarke, Jean Martin and Patrick Sturgis (2002) The Effects of Extended Interviewer Efforts on Nonresponse Bias. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 135-148.

Marker, David (2002) *Improving Quality of Surveys*. Invited paper presented at the International Conference on Improving Surveys, Copenhagen, August 2002. www.icis.dk/ICIS_papers/Keynote2_0_2.pdf

Marsh, Catherine, and Elinor Scarborough (1990) Testing Nine Hypotheses about Quota Sampling. *Journal of the Market Research Society*, Vol. 32, N° 4, pp. 485-506.

Martin, Elizabeth, Denise Abreu and Franklin Winters (2001) Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation. *Journal of Official Statistics*, Vol. 17, N° 2, pp. 267-284.

Mathiowetz, Nancy A., Mick P. Couper, Dicy Butler (2000) *The Impact of Nonresponse in the American Travel Survey*. Paper presented at the 11th International Workshop on Household Survey Nonresponse, Budapest, September 2000.

Meier, Erhard (2003) *Response Rates: The International Perspective. Existing Alternatives and their Pros and Cons*. Paper presented at the CESP symposium <Qui répond aux enquêtes> Paris, November 2003. www.cesp.org

Merkle, Daniel A., and Murray Edelman (2002) Nonresponse in Exit Polls: A Comprehensive Analysis. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 243-258.

MOA-nieuws (2003) *Infofilter verplicht voor MOA-leden*. Clou 8, december 2003, p. 22. www.uwmeningtelt.nl

Mohler, Peter, Achim Koch & Siegfried Gabler (2003) Alles Zufall oder? Ein Diskussionsbeitrag zur Qualität von face to face-Umfragen in Deutschland. *ZUMA-Nachrichten* 53, Jg. 27, November 2003, pp. 10-15.

Morton-Williams, J. (1993) *Interviewer Approaches*. Aldershot: Dartmouth Publishing.

Mowen, J.C., and R.B. Cialdini (1980) On Implementing the Door-in-the-Face Compliance Technique in a Business Context. *Journal of Marketing Research* 17, pp. 253-258.

Nicolaas, Gerry, and Nina Stratford (2004) *A Plea for the Tailored Use of Respondent Incentives*. Paper presented at the RC33 Sixth International Conference on Social Science Methodology, Amsterdam, August 2004

NSD (2004) *ESS Documentation Report 2002/2003*, 5th edition. Bergen: Norwegian Social Science Data Services. ess.nsd.uib.no/2003/ESS1DataDocReport_e05.doc.

ONS (2004) *Guidelines for Measuring Statistical Quality*. London: Office for National Statistics. www.statistics.gov.uk/about/consultations/ons_consultations/downloads/Guidelines_measuringStatisticalQuality_v1.1.pdf

O'Shea, Ruth, Caroline Bryson, Roger Jowell (2003) Comparative Attitudinal Research in Europe. European Social Survey Deliverable 1. www.europeansocialsurvey.org (ESSDOCS).

Parkinson, C. Northcote (1958) *Parkinson's Law: The Pursuit of Progress.* London, John Murray.

Pääkkönen, Hannu (1999) Are Busy People Under- or Over-represented in National Time Budget Surveys? *Loisirs & Société*, Vol. 21, n° 2, Fall 1998, pp. 573-582.

Peters, Lisa (2001) *Respondent Panel and Face-to-face Interview Compared in 'Cultural Changes in the Netherlands Survey 2000'*, The Hague, Social and Cultural Planning Office. www.scp.nl/onderzoek/respondent-panel-facetoface-interview-compared.Pdf

Philippens, Michel, Geert Loosveldt, Ineke Stoop and Jaak Billiet (2003) *Noncontact rates and interviewer calling strategies in the* ESS. Paper presented at the 14th International Workshop on Household Survey Nonresponse, Leuven, September 2003. ess.nsd.uib.no/2003_documentation.jsp (Data Quality)

Pickery, Jan, Geert Loosveldt and Ann Carton (2001) The effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Surveys, A multilevel Approach. *Sociological Methods and Research*, Vol. 29 N° 4, May 2001, pp. 509-523.

Platek, Richard, and Carl-Erik Särndal (2001) Can a Statistician Deliver, *Journal of Official Statistics*, Vol. 17, N° 1, pp. 1-20.

Pondman, L.M. (1998) *The Influence of the Interviewer on the Refusal Rate in Telephone Surveys*. Deventer: Print Partners Ipskamp.

Pondman, L.M. (1999) Overstag?! Reductie van het weigerpercentage in telefonisch onderzoek. In: A.E. Bronner et al. (eds.) *Recente ontwikkelingen in het marktonderzoek*. Jaarboek van de Nederlandse Vereniging voor Marktonderzoek en Informatiemanagement. Haarlem: De Vrieseborch, pp. 25-34.

Pring, David (2005) Will the last respondent please turn off the Internet? *Research World*, January 2005, p. 14.

Purdon, Susan, Pamela Campanelli and Patrick Sturgis (1999) Interviewers' Calling Strategies on Face-to-Face Interview Survey. *Journal of Official Statistics*, Vol. 15, N° 2, pp. 199-216.

Quintano, Claudio, and Rosalia Castellano (2001) *Strategies for Dealing with Non-responses for Quality in some Istat Surveys*. Essays, n° 11. Rome: Istat.

Reuband, Karl-Heinz, & Jörg Blasius (1996) Face-to-face-, telefonische und postalische Befragungen: Ausschöpfungsquoten und Antwortmuster in einer Großstadt-Studie. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 48, pp. 296-318.

Robinson, John P. (1999) Activity Patterns and Time-Diary Dropouts. *Loisirs & Société*, Vol. 21, n° 2, Fall 1998, pp. 551-554.

Rogelberg, Steven G., James M. Conway, Matthew E. Sederburg, Christiane Spitzmüller, Shanaz Aziz, William E. Knight (2003) Profiling Active and Passive Nonrespondents to an Organizational Survey, *Journal of Applied Psychology*, Vol. 88, N° 6, pp. 1104–1114.

Rogelberg, Steven G., Gwenith G. Fisher, Douglas C. Maynard, Milton D. Hakel and Michael Horvath (2001) Attitudes Toward Surveys: Development of a Measure and Its Relationship to Respondent Behavior. *Organizational Research Methods*, Vol. 4 N° 1, pp. 3-25.

Schlegelmilch, Bodo B., and Adamantios Diamantopoulos (1991) Prenotification and Mail Survey Response Rates: A Quantitative Integration of the Literature. *Journal of the Market Research Society*, Vol. 33, pp. 243-255.

Schmeets, Hans, and Jacques P.G. Janssen (2002) Using National Registrations to Correct for Selective Non-response: Political preference of Ethnic Groups. *Proceedings of Statistics Canada Symposium 2001*. Achieving Data Quality in a Statistical Agency: A Methodological Perspective.
www.statcan.ca/english/conferences/symposium2002/session12/s12b.pdf

Schmeets, Hans, & John Michiels (2003) Het effect van non-respons onder allochtonen. *Bevolkingstrends*, 4e kwartaal 2003, Centraal Bureau voor de Statistiek.

Schnell, Rainer (1997) *Nonresponse in Bevölkerungsumfragen, Ausmaß, Entwicklung und Ursachen*. Opladen: Leske und Budrich.

Schober, Michael F., and Frederick G. Conrad (1997) Does Conversational Interviewing Reduce Survey Measurement Error? *Public Opinion Quarterly*, Vol. 61, pp. 576-602.

Schoen, E.D., P.R Defize & M. Bakker (2000) Methodologische evaluatie van de Politiemonitor Bevolking. TNO-rapport FSP-RPT-000032. Delft: TNO-TPD.

Schonlau, Matthias, Ronald D. Fricker, Jr., Marc N. Elliott (2001) *Conducting Research Surveys via E-mail and the Web*. Rand Documents. www.rand.org/publications/MR/MR1480

Schouten, Barry (2003) Adjustment for Bias in the Integrated Survey on Household Living Conditions (POLS) 1998. Paper presented at the 14th International Workshop on Household Survey Nonresponse, Leuven, September 2003.

Schouten, Barry, Jelke Bethlehem and Ineke Stoop (2002) *Respondents, persuadable non-respondents and persistent non-respondents ; a comparison*. Paper presented at the13th International Workshop on Household Survey Nonresponse, Copenhagen, August 2002.

Singer, Eleanor (2002) The Use of Incentives to Reduce Nonresponse in Household Surveys. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 163-177.

Singer, Eleanor, Robert M. Groves & Amy D. Corning (1999) Differential Incentives. Beliefs about practices, perceptions of equity, and effects on survey participation. *Public Opinion Quarterly*, Vol. 63, pp. 251-260.

Singer, E., N.A. Mathiowetz and M.P, Couper (1993) The Impact of Privacy and Confidentiality Concerns on Survey Participation: The Case of the 1990 Census. *Public Opinion Quarterly*, Vol. 57, pp. 465-482.

Singer, Eleanor, John Van Hoewyk and Mary P. Maher (1998) Does the Payment of Incentives Create Expectation Effects? *Public Opinion Quarterly*, Vol. 62, pp. 152-164.

Singer, Eleanor, John Van Hoewyk, Nancy Gebler, Trivellore Raghunathan and Katherine McGonagle (1999) The Effects of Incentives on Response Rates in Interviewer-Mediated Surveys. *Journal of Official Statistics*, Vol. 15, N° 2, pp. 199-216.

Singer, Eleanor, John van Hoewyk and Randall J. Neugebauer (2003) Attitudes and Behavior. The Impact of Privacy and Confidentiality Concerns on Participation in the 2000 Census. *Public Opinion Quarterly*, Vol. 67, pp. 368-384.

Smeets, Ingrid (1995) Facing another gap: An exploration of the discrepancies between voting turnout in survey research and official statistics. *Acta Politica* 30, pp. 307-334.

Smith, Tom W. (1983) The Hidden 25 Percent: An Analysis of Nonresponse on the 1980 General Social Survey. *Public Opinion Quarterly*, Vol. 47, pp. 386-404.

Smith, T.W. (1984) Estimating Nonresponse Bias with Temporary Refusals. *Sociological Perspectives*, Vol. 27, N° 4, pp. 473-489.

Smith, T.W. (1995) Trends in non-response rates. *International Journal of Public Opinion Research*, Research Notes, Vol. 7, N° 2, pp. 157-171.

Smith, Tom. W. (1999) *Developing Nonresponse Standards.* Paper presented at the International Conference on Survey Nonresponse. Portland, Oregon, October 28-31, 1999.
www.norc.uchicago.edu/online/nonre.htm

Smith, Tom. W. (2002) Developing Nonresponse Standards. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 27-40.

Snijkers, Ger, Joop Hox and Edith de Leeuw (1999) Interviewers' Tactics for Fighting Survey Nonresponse. *Journal of Official Statistics*, Vol. 15, N° 2, pp. 185-198.

Snijkers, Ger, and Martin Luppes (2000) The best of two worlds: total design method and new Kontiv design. An operational model to improve respondent co-operation. *Netherlands Official Statistics*, Vol. 3, pp. 4-10.
www.cbs.nl/nl/publicaties/publicaties/algemeen/a-125/2000/nos-00-4.pdf

Spit, Jan (1996) *Onderzoeksverantwoording* AVO'95, Rijswijk: Sociaal en Cultureel Planbureau.
www.scp.nl/boeken/werkdocumenten/doc37/nl/acrobat/default.htm

SPSS (2002) *Syntax Reference Guide SPSS 11.5*. Chicago: SPSS Inc.

Steeh, Charlotte, Nicole Kirgis, Brian Cannon and Jeff DeWitt (2001) Are They Really as Bad as They Seem? Nonresponse Rates at the End of the Twentieth Century. *Journal of Official Statistics*, Vol. 17, N° 2, pp. 227-247.

Stinchcombe, Arthur L., Calvin Jones and Paul Sheatsley (1981) Nonresponse Bias for Attitude Questions. *Public Opinion Quarterly*, Vol. 45, pp. 359-375.

Stocké, Volker, and Bettina Langfeldt (2004) Effects of Survey Experience on Respondents' Attitude towards Surveys. *Bulletin de Méthodologie Sociologique*, January 2004, n. 81, pp. 5-32.

Stoop, Ineke (2000) Covering the Population, Sample Surveys at the Social and Cultural Planning Office. *Kwantitatieve Methoden*, jr. 21, nr. 65, pp. 69-84. www.scp.nl/onderzoek/km_stoop.pdf

Stoop, I.A.L. (2002a) Wie doen niet mee? Een analyse van nonrespons bij een onderzoek naar voorzieningengebruik. In: A.E. Bronner et al. (eds.) *Ontwikkelingen in het marktonderzoek*. Jaarboek 2002 Marktonderzoek Associatie. Haarlem: De Vrieseborch, pp. 69-92.

Stoop, Ineke (2002b) *Unlisted Telephone Numbers and Face-to-face Household Surveys: A Double Whammy.* Paper presented at the13th International Workshop on Household Survey Nonresponse, Copenhagen, August 2002.

Stoop, Ineke A.L. (2004a) Surveying Nonrespondents. *Field Methods*, Vol. 16, nr. 1, pp. 23-54.

Stoop, Ineke A.L. (2004b) *Nonresponse and nonprobability sampling*. Paper presented at the 15th International Workshop on Household Survey Nonresponse, Maastricht, August 2004.

Stoop, Ineke, Silke Devacht, Jaak Billiet, Geert Loosveldt and Michel Philippens (2003) *The development of a uniform contact description form in the ESS*. Paper presented at the 14th International Workshop on Household Survey Nonresponse, Leuven, September 2003.

Stoop, Ineke A.L. and Frans Louwen (1999) *From a Silent Majority to an Articulate Majority.* Paper presented at the International Conference on Survey Nonresponse. Portland, Oregon, October 1999.

Stoop, I., & J. Oudhof (1988) *Methoden van tijdsbestedingsonderzoek. Verslag van een proefonderzoek.* Paper presented at the Vlaams-Nederlandse studiedagen voor sociologen en antropologen, Antwerpen, April 1988.

Sturgis, P., and P. Campanelli (1998) The Scope for Reducing Refusals in Household Surveys: An Investigation based on Transcripts of Tape-recorded Doorstep Interactions. *Journal of the Market Research Society* 40(2) www.soc.surrey.ac.uk/~scs1ps/papers/JMRS.pdf

Sudman, Seymour (1966) Probability Sampling with Quotas. *Journal of the American Statistical Association*, Vol. 61, pp. 749-771.

Sudman, Seymour, and Graham Kalton (1986) New Developments in the Sampling of Special Populations. *Annual Review of Sociology*, Vol. 12, pp. 401-429.

Te Riele, Saskia (2002) Vertekening door non-respons. Hoe nauwkeurig zijn de uitkomsten van persoonsenquêtes? *Sociaal-economische Maandstatistiek*, Jrg. 19, April 2002. Voorburg/Heerlen: Centraal Bureau voor de Statistiek, pp. 20-25. www.cbs.nl/nl/publicaties/artikelen/algemeen/non-respons/non-respons.pdf

Teitler, Julien O., Nancy E. Reichman and Susan Sprachman (2003) Costs and Benefits of Improving Response Rates for a Hard-to-reach Population. *Public Opinion Quarterly*, Vol. 67, pp. 126-138.

Tourangeau, Roger, and Tom W. Smith (1996) Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, Vol. 60, pp. 275-304.

Tweede Kamer der Staten Generaal (1996) Aanhangsel van de handelingen, vergaderjaar 1995-1996, pp. 1853-1854. parlando.sdu.nl (kvr2863.pdf)

Väisänen, Paavo (2002) *Diary Nonresponse in the Finnish Time Use Survey*. Paper presented at the13th International Workshop on Household Survey Nonresponse, Copenhagen, August 2002.

Van Delft, Dirk (2000) Liever niet. Sociaal en Cultureel Planbureau brengt enquête-weigeraars in kaart. *NRC-Handelsblad*, 10 juni 2000.

Van den Berg, Gerard J., Maarten Lindeboom and Peter J. Dolton (2004) *Survey Non-Response and Unemployment Duration*. Tinbergen Institute Discussion Paper TI 2004-094/3 www.tinbergen.nl/discussionpapers/04094.pdf

Van der Kolk, H. (2001) Meten wat we allemaal al weten. In: A.E. Bronner et al. (eds.) *Ontwikkelingen in het marktonderzoek*. Jaarboek 2001 Marktonderzoek Associatie. Haarlem: De Vrieseborch, pp. 151-168.

Van der Wouden, Ries, & Erica de Bruijne (2001) *De stad in de omtrek*. Den Haag: Sociaal en Cultureel Planbureau. www.scp.nl/boeken/titels/2001-13/nl/metainfo.htm

Van Goor, Annemiek, & Henk van Goor (2003) De bruikbaarheid van de Methode van de Centrale Vraag voor het vaststellen van non-responsvertekeningen: vier empirische tests. *Sociologische Gids* 2003/4 (jg 50) pp. 378-391.

Van Goor, Henk (2003) Toch een groeiende kloof tussen kiezers en gekozenen? Non-respons in het Nationaal Kiezersonderzoek revisited. *Sociologische Gids* 2003/3 (jg 50) pp. 231-235.

Van Goor, Henk, and Sonja Rispens (2004) A Middle Class Image of Society. A Study of Undercovergae and Nonresponse Bias in a Telephone Survey. *Quality and Quantity*, Vol. 38, pp. 35-49.

Van Leest, Joel, & Wim Burhenne (1997) *Respons en non-respons in het Aanvullend Voorzieningengebruikonderzoek (AVO)*, Werkdocument 42, Rijswijk: Sociaal en Cultureel Planbureau. www.scp.nl/boeken/werkdocumenten/doc42/nl/acrobat/default.htm

Van Tulder, J.J.M. (1977) Op de grens van non-response. *Jaarboek van de Nederlandse Vereniging van Marktonderzoekers* 1977, pp. 43-52.

Van Tulder, J.J.M. (1978) Op de grens van non-response (2). *Jaarboek van de Nederlandse Vereniging van Marktonderzoekers* 1978, pp. 71-75.

Vehovar, Vasja, Zenel Batagelj, Katja Lozar Manfreda and Metka Zalatel (2002) Nonresponse in Web Surveys. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 229-242.

Visscher, Gerard (1995) *Kiezersonderzoek op een dwaalspoor*. Den Haag: Sdu Uitgeverij.

Voogt, Robert (2004) *"I'm Not Interested". Nonresponse bias, response bias and stimulus effects in election research*. Academisch Proefschrift. Amsterdam: Universiteit van Amsterdam.

Voogt, R.J.J., W.E. Saris and B. Niemöller (1998) Non-Response, and the Gulf Between the Public and the Politicians. *Acta Politica*, 33, pp. 250-280.

Vuijsje, Herman (1992) *Mens erger je niet: privacy en wetenschappelijk onderzoek*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
Ward, James C., Bertram Russick and William Rudelius (1985) A Test of Reducing Callbacks and Not-At-Home Bias in Personal Interviews by Weighting At-Home Respondents. *Journal of Marketing Research*, Vol. XXII (February 1985), pp. 66-73.
Warriner, Keith, John Goyder, Heidi Gjertsen, Paula Hohner and Kathleen McSpurren (1996) Charities, no, lotteries, no, cash, yes: Main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly*, Vol. 60, pp. 542-562.
Werkgroep dataverzameling (1996) *Zicht op nonrespons?* Werkdocumenten werkgroep nonrespons. Vereniging voor beleidsonderzoek (VBO) Ubbergen: Tandem Felix.
Willimack, Diane K., Howard Schuman, Beth-Ellen Pennell and James M. Lepkowski (1995) Effects of a Prepaid Nonmonetary Incentive on Response Rates and Response Quality in a Face-to-face Survey. *Public Opinion Quarterly*, Vol. 59, pp. 78-92.
Zuzanek, Jiri (1999) Non-response in Time Use Surveys: Do the Two Ends Meet? *Loisirs & Société*, Vol. 21, N° 2, Fall 1998, pp. 547-549.

Curriculum vitae Ineke Stoop

Born in The Hague, the Netherlands on 14 March 1953

Education

1965 - 1971	Secondary school: gymnasium ß, Scholengemeenschap Simon Stevin, The Hague
1971 - 1980	Master's study Psychology, Leiden University, master's degree (cum laude) main subject methodology of psychological research, subsidiaries social psychology and data theory

Employment

1973 - 1980	Research assistant Leiden University
1980 - 1983	Researcher department of Data Theory, Leiden University
1983 - 1990	Staff member department of Socio-cultural Statistics, Statistics Netherlands
1990 - date	Head of department of Data Services and IT, Social and Cultural Planning Office of the Netherlands

Professional memberships

Elected member of the International Statistical Institute (ISI)
Member of The World Association of Research Professionals (ESOMAR)
Member of the MarktOnderzoekAssociatie (MOA)
Member of the Netherlands Society for Statistics and Operations Research (VVS)
Member of the Dutch/Flemish Classification Society (VOC)

Other appointments and professional activities

Member of the Central Coordinating Team (CCT) of the European Social Survey (ESS)
Member of the European Advisory Committee on Statistical Information in the Economic and Social Spheres (CEIES)
Member of the Statistics Netherlands Advisory Council on Methodology and ICT
Deputy member of the Appeal Commission Code of Conduct Research and Statistics MOA, VBO, VSO
Former board member of the Scientific Statistical Agency (WSA) of the Netherlands Organisation for Scientific Research (NWO)